

Interstitial solution enthalpies derived from first-principles: Knowledge Discovery using High-Throughput Databases

Ugur Aydin

Paderborn 2016

Interstitial solution enthalpies derived from first-principles: Knowledge Discovery using High-Throughput Databases

Dissertation
zur Erlangung des akademischen Grades
Doktor der Naturwissenschaften (Dr. rer. nat.)
vorgelegt dem
Department Physik der Fakultät für Naturwissenschaften
an der Universität Paderborn

Ugur Aydin

Promotionskommission

Vorsitzender	Prof. Dr. Dirk Reuter
Gutachter	Prof. Dr. rer. nat. Wolf Gero Schmidt
Gutachter	Prof. Dr. rer. nat. Jörg Neugebauer

Tag der Einreichung:	1. März, 2016
Tag der Verteidigung:	6. April, 2016

Abstract

The incorporation of interstitial elements in transition metals can drastically change the performance of these materials. For example, the incorporation of hydrogen, which is often unavoidable, can cause devastating failure due to embrittlement. Interstitial carbon and nitrogen for instance favor the formation of microstructures (e.g. carbides or nitrides). Boron can stabilize grain boundaries and oxygen causes unwanted material oxidation. In each of these cases, the knowledge of the atomic solution enthalpy is of enormous importance, since it determines the concentration of the inclusions.

In this thesis, a comprehensive ab initio study of solution enthalpies of typical interstitial elements H, B, C, N, O, F, (He and Ne) in 3d, 4d and 5d transition metals is performed. The primary objective of the analysis are mechanical and chemical regularities and trends for the solubility in order to improve the prediction of atomic concentration in a variety of materials. One challenge is to identify quickly accessible physical material parameters, so-called descriptors that depend only on the host lattice and may allow to predict the solution enthalpy of interstitial elements in materials not considered in this study. In order to determine material parameters such as band structures, atomic positions, and electronic density of states, ab initio methods within the density functional theory (DFT) framework were used, which allow the calculation of the energetics of a material on the electronic scale. The detection of trends is achieved through the generation and analysis of large amounts of ab initio data. To handle such amount of data, a Python-based working environment for data management (generation, storage, visualization, and analysis) was developed. Tools of multivariate statistics and numerical analysis (for example factor analysis or correlation coefficients) have been implemented and used for the analysis.

In a multistage process, first the solution enthalpy of hydrogen in the 3d transition metals has been studied extensively. An universal trend for the solution enthalpy in dependence of a chemical and a mechanical effect could be observed. The study has been extended to other interstitial elements and transition metals. In order to analyze the large amount of produced data we used the factor analysis and found that all interstitial element solution enthalpies can be described by three common yet unknown factors. On closer inspection we found that boron is an outlier with one dominant factor and its exclusion yields a description of the solution enthalpy for which two factors are sufficient. Furthermore, the classification of the interstitial elements into two groups revealed that only one common factor can be used. Specifically, for the elements H, C, N, and O the factor can be related to the atomic distance and the electronic density of states of the host lattice which allows the prediction of the solution enthalpy with high accuracy. The insights provided in this work are useful guidelines in materials alloy design and the identified descriptors can be employed for screening candidate materials in existing structural databases.

Contents

1	Introduction	9
2	High-Throughput in Materials Science	13
2.1	Automatized experiments	13
2.2	Computational approach	16
3	Theoretical Background	19
3.1	From the many-body problem to Density Functional Theory	21
3.2	Solubility of interstitial elements	28
3.3	Ab initio description of the solution enthalpy	30
3.4	Zero point energy: Einstein method and dynamical matrix	31
3.5	Semi-empirical models for interstitial solubilities	33
3.6	Bravais-Pearson correlation coefficient	37
3.7	Factor Analysis	39
4	The workbench	41
4.1	Concepts	41
4.2	Implementation	43
4.3	Tools of the workbench	49
5	Results I: Study of chemical trends	54
5.1	Study of host materials	54
5.1.1	Self-consistent parameter optimization	54
5.1.2	Chemical trends for host materials	56
5.2	Comparison of hydrides, borides, carbides, and nitrides	61
5.3	Study of hydrogen in 4th row elements	63
5.3.1	Comparison of solution enthalpies with experiments	63
5.3.2	Solution enthalpy as a function of the lattice constant	65
5.3.3	Hard spheres model	65
5.3.4	Separation of strain and chemical effects	68
5.3.5	Zero point vibrations	70
5.3.6	Chemical shift of ΔH	70
5.3.7	Comparison with semi-empirical models	72
5.4	Hydrogen solution enthalpy in 5th and 6th row elements	73

5.4.1	Hydrogen solution enthalpy in ground state structure	73
5.5	Study of interstitial elements in transition metals	76
5.5.1	Chemical potential of the interstitial elements	76
5.5.2	Interstitial solution enthalpy in NM fcc structure - d-band trend	77
5.5.3	Interstitial ΔH vs B_0 and ΔE_{F-s}	79
5.5.4	Electronic effects and charge density differences	82
6	Results II: Knowledge Discovery in Databases	85
6.1	Correlations between solution enthalpies of interstitial elements	85
6.1.1	Linear dependence between solution enthalpies	86
6.1.2	Chemical vs elastic effect on the interstitial elements	88
6.1.3	Sharpening chemical trends	89
6.2	Factor analysis of the solution enthalpies	90
6.2.1	Determination of the number of factors k	91
6.2.2	Factor analysis for NM fcc structure	93
6.2.3	Search for the best ΔH predicting equation and corresponding bulk parameters	93
6.2.4	FA for physical bulk properties	97
6.2.5	Results for the combinatoric search	97
6.2.6	Conclusions from the combinatoric search	101
6.2.7	Master curve for interstitial solubility	104
7	Summary and Outlook	106
A	Correlation of Griessen's predictive equation	110
B	Study of interstitial elements	111
B.1	Hydrogen solution enthalpy vs lattice constant	112
B.2	Derivative of hydrogen solution enthalpy	113
B.3	Interstitial solution enthalpies vs lattice constant	114
B.3.1	Boron solution enthalpy vs lattice constant	114
B.3.2	Carbon solution enthalpy vs lattice constant	115
B.3.3	Nitrogen solution enthalpy vs lattice constant	116
B.3.4	Oxygen solution enthalpy vs lattice constant	117
B.3.5	Fluorine solution enthalpy vs lattice constant	118
B.3.6	Helium solution enthalpy vs lattice constant	119
B.3.7	Neon solution enthalpy vs lattice constant	120
B.4	Charge density differences of interstitial elements	122
B.4.1	CDD of Interstitial elements in early TMs	122
B.4.2	CDD of Interstitial elements in half-filled TMs	123
B.4.3	CDD of Interstitial elements in late TMs	124
B.4.4	CDD of H in all TMs	125

C Factor Analysis	126
C.1 Solution enthalpy matrix	127
C.2 Standardized solution enthalpy matrix	128
C.3 Factor Analysis of ΔH for interstitial elements H, B, C, N, O, F, He, Ne	128
C.3.1 Determination of the number of factors k	128
C.4 Results/Tables for the combinatoric search	132
Bibliography	142

Chapter 1

Introduction

The design of materials by generating and analyzing large databases is at latest since the “Materials Genome Initiative” of Barack Obama a strategy that attracts more and more attention in the community [1]. The idea of high-throughput (HT) investigations is, however, much older, since first HT experiments were performed and reported in the field of materials research more than 130 years ago [2]. Some of the resulting discoveries are well-known due to the significant impact on our daily life, while much less is known about their HT background. It is worth mentioning two examples here:

In October 1878, Thomas A. Edison registered the patent “Improvement In Electric Lights”, a long lasting electrical light bulb, containing an incandescent metal-filament being a conductor with a high fusing-point. However, it took him 40,000 notes pages filled with the results of testing 1,600 different elements (even including “hears from a friend’s beard” [3]) in order to discover a suitable material for the incandescent filament. Finally he suggested an optimal material, a carbonized bamboo fiber in a vacuum light bulb which could last for approximately 1200 hours (50 days) and was registered as a patent in November 1880 [4]. It goes without saying that this discovery had for many generations a significant impact on their daily (nightly) life.

In 1912 Ciamician tested hundreds of flasks with potentially photoactive materials in order to determine the possibility of using an alternative and renewable energy form instead of the limited fossil energy (which was mainly coal at that time). He shared his vision of a clean photochemical industry instead of an industry based on coal usage for covering the energy needs. Today, more than 100 years later, this topic is one of the hottest topics in materials science. In addition, his method of spectroscopy helped Mendelejew to develop the periodic table of elements. However, one can imagine from figure 1.1 that he might have faced spacing problems for storing all the flasks on the roof of the University of Bologna in order to determine the possibility of using a photochemical process for batteries [5].

In the following, HT experiments found their way and became of central importance in chemical industry and for the development of parallel screening and parallel analysis methods in order to decrease the time of finding new drugs [7]. The same development is currently going on in the field of materials science. In the majority of these cases [8, 9], HT is used to screen large amounts of chemical compositions in a “trial and error” approach for materials that optimize certain predefined properties such as thermoelectric, magnetic or mechanical key quantities. Only in exceptional cases, like the field of computational thermodynamics [10, 11], experimental databases are used to identify

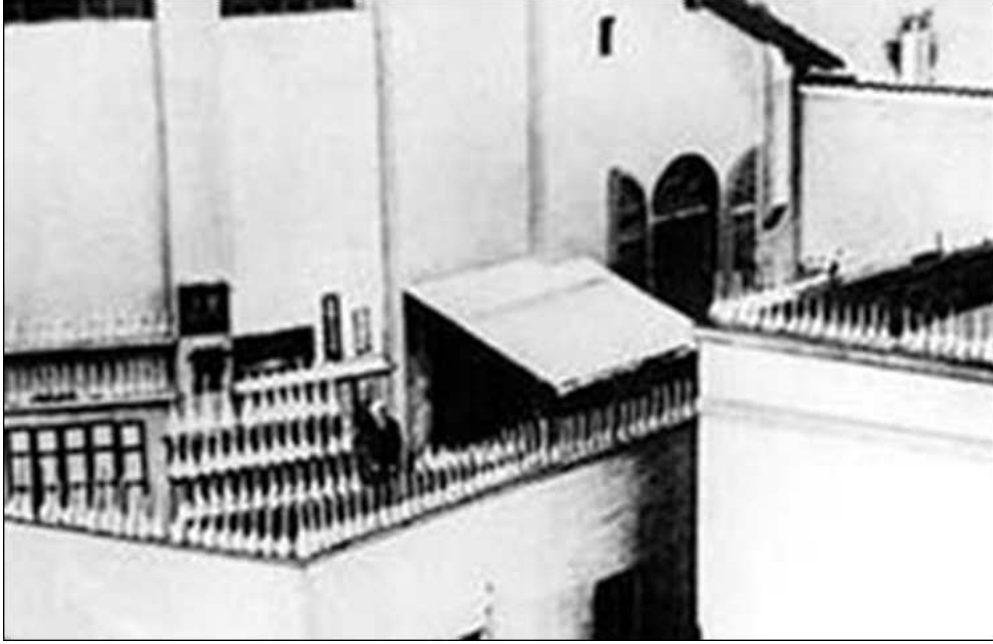


Figure 1.1: Ciamician testing hundreds of samples in parallel on the roof of the University of Bologna. Taken from [6].

generic rules for physical properties.

Besides the experimental research approaches, the increasing computer power in particular in high-performance clusters (HPC) along with tremendous developments of theoretical and numerical methods are pushing the importance of computer aided research methods since 50 years forward. The field of modern materials research is heavily affected by this development. Computational materials research nowadays covers various scales between the electronic and the macroscopic level, i.e. length scales between 10^{-10} m and 1 m and time scales between 10^{-15} s and 10^7 s. In this thesis, mainly the electronic scale is exploited and the quantum-mechanical Schrödinger equation is solved without using empirical parameters. Due to the huge amount of interacting atoms and electrons ($\approx 10^{23}$) these so called *ab initio* approaches are typically not exact, but rely on approximations. The Born-Oppenheimer approximation [12] and the development of the Density Functional Theory (DFT) [13] are the most established concepts to overcome the problem of treating a huge number of interacting particles and are therefore employed in this work.

An advantage of DFT as compared to alternative atomistic methods, is the possibility to derive chemical or structural trends based on an efficient simulation of the electronic structure. On the one hand, methods that by construction would allow more accurate results, such as the solution of the Schrödinger's equation with quantum Monte-Carlo techniques [14], have too many uncontrollable approximations. The required approximations in DFT, which make the approach much more effective, might have an influence on absolute numbers like formation energies, but the relative numbers in a trend study are typically predicted rather accurately [15]. Approaches that reduce the complexity even further, such as the application of empirical potentials [16–20] are not based on the

evaluation of the electronic structure, but on a fitting to experiments and are therefore much less predictive. Experiments, on the other hand, can typically only be performed for selected parameters in a “trial and error” approach, since the consideration of hundreds of experiments is logistically difficult or requires compromises in terms of accuracy.

The nature of *ab initio* calculations allows to modify systematically the chemical composition, crystalline structures (e.g. the lattice constant) and physical parameters (e.g. the magnetization), choosing even values that cannot be reached in nature. In this way, much broader trends of all microscopically determined physical properties such as atomic positions/relaxations, band gaps, density of states, etc. can be achieved, than possible in experimental materials design [21]. The drawback of the application of DFT with its focus on the electronic structure is, however, the difficulty to determine macroscopic properties (such as formability, embrittlement etc.) directly. The most demanding part of DFT based HT methods for materials design is, therefore, to identify promising and fast so called descriptors [22–27]. These descriptors are essential for the communication with the database and usually needed to connect the microscopic parameters mentioned above with the macroscopic properties. Such a microscopic parameter is for example the intrinsic stacking fault energy γ_{isf} which can be linked to the macroscopic deformation behavior of the material. A low γ_{isf} leads to a strengthening as shown for example for Mg-Y [28] or Fe-C alloys [29].

The field of DFT based HT investigations [30–35] is still relatively young, but is growing fast in the field of materials research. The latter can be mainly subdivided into functional and structural materials. Both classes have in common that transition metals (TMs) play the most important role to achieve a certain (for example electronic [36, 37], magnetic [38] or mechanical [39]) functionality. It is, however, clear that a perfect crystal consisting solely of one TM element can be synthesized in labs but does not exist under real conditions. They are always alloyed with other chemical species. In particular when being in contact with an (often reactive) environment, the incorporation of interstitial elements cannot be avoided, since they are typically rather mobile and therefore able to penetrate into the lattice [40]. Through several mechanisms they can have a favorable or detrimental impact on the material properties. For example they influence the stacking fault energy [29, 41], yield precipitate formation [42, 43] or influence the mobility of dislocations [44]. In this way, one of the impacts of interstitial atoms (e.g. reported for carbon [45]) is a crystal hardening or softening. Alternatively, they can lead to serious material failure like interstitial embrittlement known for example for hydrogen [16, 46–48] and nitrogen [49, 50]. One goal in this field is to gain control over these property changes. Therefore, it is very important to know the underlying mechanisms governing the interstitial solubility and to a minor extend the interstitial diffusivity within these materials.

In order to understand the mechanisms governing interstitial solubility a wide range study for different interstitial elements including H, B, C, N, O, F, He, and Ne in all TMs and neighboring host elements (in total 32 host elements) is performed in the present work. A variety of different structures (fcc, hcp, bcc), magnetic states (fm, nm, and afm) and interstitial positions are considered. Our aim in this work is not in the first place a ranking of solubilities, which would be in line with several other state-of-the-art HT studies [8, 51–54] as for example presented for hydrogen storage materials in chapter 2, but a detailed analysis of the generated data to obtain novel insights about underlying mechanisms for interstitial solubilities in TMs. These insights are very important since a general understanding of the coherency between chemical and structural solubility mechanisms

can allow one to predict interstitial solubilities in more complex host materials without performing time and money consuming experiments or calculations.

The interstitial solubility (i. e. the concentration c_H) can be determined for example with the equation [55]

$$c_H = \sqrt{\frac{p}{p_0}} \cdot e^{\frac{\Delta S}{k_B}} \cdot e^{-\frac{\Delta H}{k_B T}}, \quad (1.1)$$

in which T denotes the temperature, p is the pressure, p_0 the reference pressure (usually $p_0 = 1$ atm), k_B the Boltzmann constant, ΔS is the entropy change between the reactant and product system, which is assumed to be constant with respect to the temperature T . The last term ΔH denotes the solution enthalpy of the interstitial element which is of main focus in this thesis since it is directly linked to the solubility as shown above.

Starting with standard ab initio calculations (e.g. total energies of a large variety of systems) and their transformation to higher level data (e.g. the solution enthalpy of interstitial atoms), the data mining step employs statistical approaches in order to evaluate and find patterns/correlations between the solution enthalpy and (combined) material characteristics ("fast descriptors") of the pure host materials. Once relevant patterns are found, the newly gained knowledge for solubility trends can be used to predict interstitial solubilities for compositions not considered in the database and eventually to suggest new host materials with a tailored interstitial solubility. Our such developed Knowledge Discovery (KD) process is constructed such that it can be fully automatized and can therefore be applied to various other materials science questions.

Chapter 2

High-Throughput in Materials Science

The two examples of Edison and Ciamician indicate that setting up HT experiments by hand can be a cumbersome and error prone task. It is obvious that an automated setting up of the test-samples and their automated analysis, including an automated documentation would have increased the speed of finding the mentioned materials. Such an automatization, however, requires the availability of the technical infrastructure and is also closely related to computational facilities.

First efforts to perform automated HT experiments have therefore only started approximated 50 years ago. In 1970 Hanak introduced a “new parallel” approach to search and find novel binary compositions that become superconducting at sufficiently low temperatures [56]. These measurements, which will be discussed in more detail below, can be considered as a milestone in automated HT research, since they documented for the first time a thirty-fold increase of the rate of experimental output due to automatization. With the development of computers and computational power, automated HT methods have in the last two decades more and more often be performed solely on the computer. The development of technologies such as databases, data-mining technologies, modeling approaches, and evolutionary strategies allowed one to further speed up the discovery and development processes substantially [2, 57]. These developments are central for the present thesis, and some landmarks in the field will therefore also be discussed below.

2.1 Automatized experiments

As mentioned above, Hanak suggested in 1970 the key concept for a new approach in materials design, which contains three parallel steps:

- the parallel synthesis of an entire multicomponent system in a single experiment,
- the parallel chemical analysis of the entire system based on a few simple measurements, and
- the materials testing and evaluation in a single experiment.

The keys to his success was the simplification of a one-cathode, multiple-target, radio-frequency co-sputtering synthesis technique and the development of a computerized analytic method to measure a simple and common property of the sample (the thickness of the film). It allowed the implementation of a so called “multiple-sample concept” in which binary and ternary solid alloy systems are synthesized and their analysis is performed in a single experiment.

In his setup, two different target materials A and B are placed as half-discs on the “radio-frequency” cathode within a chamber filled with constant argon gas. Ionized argon atoms simultaneously evaporate atoms from this target and result into a “co-sputtering” of A and B on a single water-cooled substrate such that almost the entire composition continuum of the binary system is captured.

For three-component thin films the same technique is introduced. The second key invention was the development of a parallel chemical analysis by calibrating the sputtering apparatus in order to measure the film thickness. From the measured film-thickness and assuming Vegard’s law [58] the element composition is obtained. When Hanak applied this technique to the determination of superconducting materials, 50 gold contacts were evaporated on such a thin film in order to measure the transition temperature and the resistivity. With this approach Hanak achieved a thirty times higher rate of experimental output for binary systems and has paved the way for presently used high-end HT materials design.

The improvement of this kind of synthesis methods such as pooled synthesis (e.g., split-pool) and masking strategies (e.g., gradient or x-y shutter masking) [59] increased the efficiency and accuracy of material library generation. These sophisticated techniques have been applied to a wide range of materials. Some of the most important materials under focus are: superconductor materials [60], ferroelectric materials [61], magnetoresistive materials [62], structural materials [63], hydrogen storage materials [8, 51–53, 64], ferromagnetic shape-memory alloys [65], various types of catalyst materials [66–68], materials for fuel and solar cell applications [69–72], and automotive coatings [73]. This list of materials and references could be extended, and it is clear that not the whole set of materials properties can be considered in the present thesis. Instead we will focus our considerations on interstitial solubility in metals. In this context, some interesting experimental studies have for example been performed on hydrogen storage materials.

Hydrogen storage materials

In 2004 Olk has presented a combinatorial approach to materials synthesis and the screening of their hydrogen storage capabilities [51]. Mg-Ni-Fe ternary and Mg-Ni or Mg-Fe binary thin-films (pads) have been synthesized, employing pulsed laser deposition as well as magnetron sputtering. The element Mg was first sputtered under argon atmosphere over the entire substrate and then the elements Ni and Fe are evaporated by using a combination of laser pulses and mask-slot positioning. In this way, a 4 by 4 mask with 16 ternary Mg-Ni-Fe pads and 32 Mg-Ni or Mg-Fe binary pads in the intersections was created. In a second chamber with H₂ atmosphere the sample is loaded with hydrogen by controlling the substrate temperature. For the subsequent analysis, infrared imaging is used in order to screen the materials. The physical background is a modified electronic density of states (DOS) and a rise of the Fermi energy E_f due to the absorption of hydrogen in intermetallic alloys. More precisely, E_f is shifted to a region with a lower DOS, which leads to a less metallic character and therefore an increase in the electrical resistance. At the same time, the materials are changing their optical properties due to the charging with hydrogen [51]. This change can be detected due to IR emissivity change. The optical approach has the advantage that a parallel analysis of 16 separate Mg-Ni-Fe ternary pads and 32 Mg-Ni (Mg-Fe) binary pads in a single measurement became possible.

In 2006 Gremaud *et al.* used a similar combinatorial method that could examine 10^3 samples in parallel. Again, the metal-insulator character during hydrogen loading of the gradient sample is used. During the so called hydrogenation [54] all hydride forming compositions are mapped simultaneously. This method, called hydrogenography, has been successfully applied to $\text{Mg}_x\text{Al}_{1-x}\text{H}_y$ thin films [53]. One year later Gremaud and co-worker could successfully adapt hydrogenography to metallic hydrides such as Pd-H, which is known not to experience a metal-insulator transition [52]. The “pressure-optical-transmission-isotherms” (PTI) of metallic hydrides is used, i.e. the optical transmission of diffuse light during the hydrogenation of the substrate, which correlates linearly with the hydrogen concentration. Furthermore, the hydrogen enthalpy and the entropy could be obtained.

In 2007 Ludwig *et al.* [8] used a three level substrate as shown in figure 2.1 instead of sapphire as a substrate. Here, the substrate is a cantilever which is bending during hydrogenation of the

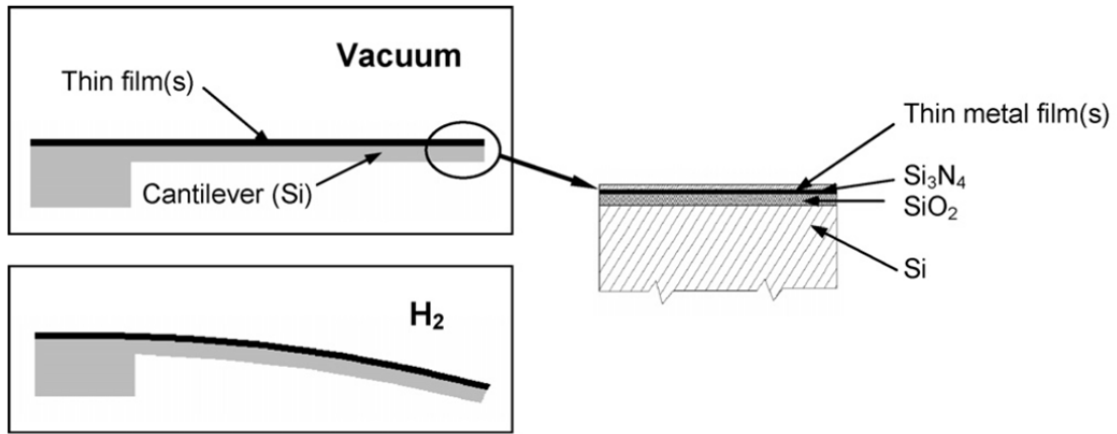


Figure 2.1: Figure copied from [8]. Three level substrate layers (micro machined Si cantilevers, SiO₂ and Si₃N₄) used by Ludwig *et al.*. Hydrogenation of the thin films and its stress change leads to a bending of the film/cantilever.

thin film. The volume expansion of the thin film leads to an out-of-plane strain and the in-plane mechanical stress obtained from bending is directly linked to the H concentration. Their concept allowed to measure 16 film/cantilever in parallel.

It is an advantage of computer aided HT DFT approaches that they do not depend on the existence of such complex setups or empirical laws for the interpretation of the results. One should note, however, that even in the case of DFT calculations, the automatized setup of structures and chemical configurations is a non-trivial task. This is currently under development and was a major part of the present PhD thesis. However, it should be mentioned here that the accuracy of DFT calculations are limited due to intrinsic (e.g. the choice of the exchange-correlation) and numerical errors (e.g. numerical convergence). Using only DFT calculations, which is an approach at $T = 0$ K to describe/predict realistic material behavior at finite temperatures can fail since further effects contribute to the free energy such as anharmonic, magnetic excitations, and electronic effects. They are, however, of little importance for the total energy at 0K (which is the temperature considered in this work), while they will become more dominant with increasing temperature.

2.2 Computational approach

Computer aided HT materials research based on DFT calculations can replace expensive “trial and error” experiments. Several advantages of the ab initio character of DFT become immediately apparent: One is not confronted with experimental constraints such as the impact of substrates or reactive environments. One can easily adjust the desired crystalline structure without inventing complicated synthesis routes. One has full control over the chemical composition and can avoid impurities. One can adjust physical parameters such as temperature, pressure, or magnetization. Moreover, a computer-based approach is ideally suited for automatizations in which certain characteristics of materials (structure, chemistry or physics) are systematically modified.

According to Curtarolo *et al.* [22] a computational HT method should be subdivided into the three connected steps

- (i) virtual material growth,
- (ii) rational materials storage, and
- (iii) materials characterization and selection.

Step (i) corresponds to setting up the structure and performing subsequent thermodynamic and electronic structure calculations of materials. Step (ii) is the storage of all relevant information in databases (DB). Examples for such DBs and corresponding tools are

- Open Quantum Material Database (OQMD) [35],
- Materials Project [33],
- Computational Materials Repository (CMR) [74],
- AFLOWLIB [75],
- Pearson’s crystal data (PCD).

OQMD is a database containing DFT calculated properties of 285,780 compounds. Materials Project is an open web-based computed information on known materials for 65,968 compounds. CMR is a software infrastructure which provides tools for data management (storing, analyzing, etc.) with the main focus on DFT calculations. Up to now 126,569 data entries are available on their web page. Aflowlib is a globally available DFT based database of 771,310 material compounds. PCD is a crystal structure database for Inorganic Compounds containing 274,000 structural data sets.

The presumably most demanding step (iii) is the data analysis for the selection of novel materials according to user-defined criteria or the discovery of new physical insights. The authors distinguish between intrinsic and extrinsic HT computational research. The former dominates the current literature in this field. It consists only of the characterization and selection step (iii) and is performed through (fast) descriptors, which are empirical quantities. These descriptors form a link between microscopic data and macroscopic physical parameters and are mandatory to determine the optimum material within an existing DB.

Extrinsic research is a closed feedback loop over the steps (i), (ii) and (iii). The evaluation of descriptors on existing DBs will suggest new structures, which will generate further data for the

DBs and so on. The suggestion of new structures can be obtained from different schemes such as genetic algorithms [76, 77], data mining [78], bayesian probabilities [79], or neural networks [80, 81]. Since these strategies are not subject of this thesis, they are not discussed here.

More important for the present study is the fact that also in the extrinsic approach, the identification of accurate and fast descriptors is a bottleneck for an efficient communication with the DB. For certain applications such a descriptor is known from long-standing empirical studies. In the field of structural materials, for example, the above mentioned stacking fault energy is known to be a key parameter that fulfills two aspects: On the one hand, it controls macroscopic deformation mechanisms such as twinning or transformation induced plasticity (TWIP and TRIP) that mesoscopically involve a change of the stacking sequence of atomic layers. On the other hand, it is relatively quickly accessible by ab initio calculations. This descriptor is for example used to optimize high-strength steels or Mg alloys with the help of DFT calculations [82–84].

In more sophisticated approaches, however, the explicit form of the descriptor is a priori not known, but this is part of the evaluation process. For example, Yang *et al.* [85] have used HT ab initio methods to determine candidates for Topological Insulators (TI), where a band inversion is induced by spin-orbit coupling. The investigations started, with a DFT study of the energy difference of conduction and valence bands, were extended to its strain dependence and finally considered the ratio of DFT calculations with and without spin-orbit coupling

$$\chi_{\text{TI}} = -\frac{E_k^{\text{SOC}}(a_0)/a_0}{\delta E_k^{\text{noSOC}}(a)/\delta(a)|_{a_0}}. \quad (2.1)$$

Once the descriptor parameter χ_{TI} has been identified, it can be used to get a fast feedback for the feasibility of all materials in a repository. In the present case [85], the repository *aflowlib.org* with the mentioned descriptor led to the discovery of 28 (already known and also unknown) TIs.

In general, the descriptor can contain more than one physical quantity. In the of field thermoelectric materials, for example, such a descriptor is the dimensionless figure of merit ZT , which combines the electrical and thermal conductivity (σ and κ) such that the equation

$$ZT = \frac{S^2 \sigma T}{\kappa} \quad (2.2)$$

in which S corresponds to the Seebeck coefficient and T to the temperature, can be used in order to quantify the thermoelectric performance. The material properties σ , κ are evaluated from the ab initio band structure. In order to determine thermoelectric materials ZT must be at least larger than one. A stability descriptor under focus is the enthalpy of formation, which is also accessible with ab initio methods. Both descriptors were used in different studies of Madsen *et al.* [86–88].

In 2016 Kirklin *et al.* used HT DFT computational approach to search for promising precipitate/alloy system. For this purpose they define a four compound descriptor, i.e. Stability, Coherency, Cost, and Two-phase equilibrium with the host [89]. 200,532 DFT calculations (L1₂, D0₁₉, L2₁ structures) have been performed in order to search for strengthening precipitates in alloy matrices. First the convex hull (it represents the most stable structure with the lowest energy for the given compound) from the OQMD is determined for all considered compounds. The stability (the first descriptor) is given as the difference energy between the phase under focus and the energy of the convex hull at the composition for that phase taken from the OQMD. All compounds with

a difference larger than 25 meV/atom are ignored. The coherency is defined as a relative lattice mismatch between the precipitate compound and the host element. All compounds exceeding the mismatch of 10 percent are ignored. The cost is determined from the so called Herfindahl-Hirschman Index (HHI), which is an availability measurement, as used by Gaultois *et al.* [90]. The last descriptor, the so called two-phase equilibrium with the host is valid if the precipitate compound is stable or if it at least decompose into the bulk.

The simultaneous relevance of several physical parameters is occasionally visualized in so-called treasure maps. For example such a treasure map was presented by Lencer *et al.* for phase-change materials [91]. Materials with a small degree of ionicity and small magnitude of covalency of the bonding have the tendency towards phase-change. In 2013 Hammerschmidt *et al.* demonstrated that the composition-dependent relative volume difference and the valence electron count employ the structural stability of TCP phases in multi-component TM alloys [92]. Koermann and co-workers [93] presented the Curie temperature and magnetization saturation (at $T = 0$ K) maps for CoCrFeNi-based alloys w.r.t. (Ag,Au,Pd,and Au) and Cr content.

However, the consideration of different physical parameters in one descriptor is typically not straight-forward. It is neither clear, which physical parameters should be involved nor how they should be combined in a mathematical expression. This challenge can be addressed by multivariate statistic methods, such as Multi-optimization techniques, Factor Analysis (FA) or Principle-component analysis (PCA). The application of these approaches implies that the generated set of data is used to generate a knowledge about physical relations and mechanisms, which is not available at the beginning of the study. Such a knowledge discovery is performed in the present thesis by using the FA which will be described below.

Chapter 3

Theoretical Background

The combination of HT with DFT based computer simulation(s) is a very powerful and fast method in order to systematically study and understand the mechanisms on the electronic scale governing interstitial solubilities. For transition metals it is known that electronic effects (e.g. hybridization, screening, etc.) directly influence for example the hydrogen solubility which is one out of several interstitial elements under focus in this thesis. An exact solution of the quantum mechanical problem on the electronic scale would need to consider approximately 10^{23} interacting particles (electrons and atoms) at the same time which is not possible. Simplifications of the Hamiltonian are needed and presented in the following chapter.

The energy E of a system of interacting nuclei and electrons, both in motion, is given by the time-independent Schrödinger equation.

$$\tilde{H}\psi = E\psi, \quad (3.1)$$

in which the Hamilton operator \tilde{H} denotes the sum of the kinetic operator \tilde{T} and the potential operator \tilde{V} ($\tilde{H} = \tilde{T} + \tilde{V}$). On the electronic scale the kinetic operator describes the electronic motion (atomic units are considered as $\hbar = m_e = e = 4\pi/\epsilon_0 = 1$)

$$\tilde{T}_{\text{el}} = -\frac{1}{2} \sum_{i=1}^{N_{\text{el}}} \nabla_i^2 \quad (3.2)$$

and the nuclei motion

$$\tilde{T}_{\text{nuc}} = -\frac{1}{2} \sum_{J=1}^{N_{\text{nuc}}} \frac{\nabla_J^2}{M_J} \quad (3.3)$$

with N_{el} electrons and N_{nuc} nuclei. M_J are the masses of the nuclei and ∇ the Nabla operator. If the electron and nucleus interaction is considered the potential operator \tilde{V} contains the repulsive electron-electron

$$\tilde{V}_{\text{el-el}} = \frac{1}{2} \sum_{i \neq j}^{N_{\text{el}}} \frac{1}{|r_i - r_j|}, \quad (3.4)$$

the repulsive nucleus-nucleus

$$\tilde{V}_{\text{nuc-nuc}} = \frac{1}{2} \sum_{J \neq I}^{N_{\text{nuc}}} \frac{Z_I Z_J}{|R_I - R_J|}, \quad (3.5)$$

and the attractive electron-nucleus

$$\tilde{V}_{\text{el-nuc}} = - \sum_{J=1}^{N_{\text{nuc}}} \sum_{i=1}^{N_{\text{el}}} \frac{Z_J}{|r_i - R_J|} \quad (3.6)$$

Coulomb interaction, with r_i denotes the coordinates of the i -th electron, R_I (Z_I) the coordinates (charge) of the I -th nucleus.

Even for a system with just three interacting bodies the Schrödinger equation cannot be solved analytically. Therefore a numeric solution is necessary. A quantum mechanical system with $\approx 10^{23}$ interacting bodies cannot be solved at all. Here, depending on the required accuracy, approximative simulation methods are applied in the field of Computational Materials Research for different time and length scales.

Typical length and time scales of materials simulations

- **Electronic scale simulations:** 10^{-9} m and 10^{-15} s
- **Atomic scale simulations:** $10^{-9} - 10^{-6}$ m and 10^{-9} s
- **Film scale simulations:** 10^{-6} m and $10^{-6} - 10^{-4}$ s
- **Macroscopic scale simulation:** 1 m and 10^7 s

Following this philosophy of approximative methods and their scales, we adapted three external programs, namely VASP (DFT), LAMMPS (Molecular Dynamics), and KMC (Kinetic Monte Carlo) representing the first three simulation scales into our workbench having in mind that coupling between the methods of different length scales (or intermediate scales) are also possible. A separate code for the macroscopic scale calculating material properties of multi-phase aggregates (figure 3.1) has been developed in the same programming language python [94–96]. This code is not adapted to the workbench but in principle can be easily linked to it since it is using elastic constants of single-phase crystals.

Whereas VASP is a so called ab initio method, i.e. free of empirical parameters, LAMMPS and KMC require further input. For example, in the case of LAMMPS this could be an empirical potential and a thermostat or in the case of LAMMPS or diffusion barriers and binding energies in the case of KMC. In the latter two cases electronic interactions are not considered directly in the Hamiltonian. They are contributing to the empirical potential of LAMMPS and to the binding energies and diffusion barriers of KMC. The needed input information can be gained by VASP. The implemented workbench allows the communication between the different length scales. For example, the output of VASP calculations can be used as input parameters for the larger scales tools and vice versa. This might be a fast relaxed structure performed with LAMMPS as a direct input structure for the computationally more demanding VASP calculations.

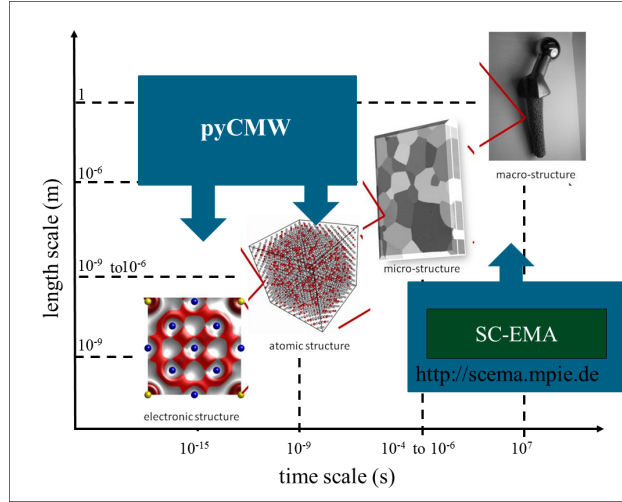


Figure 3.1: Different time and length scales covered by pyCMW and SC-EMA.

In all three scales, if one looks at the huge amount of $\approx 10^{23}$ atoms (plus electrons for electronic scale simulations) in a crystal, it becomes clear that there is a huge number of interactions which needs to be considered.

Even for a single atom, we take, for example, Fe, contains 26 electrons plus one nucleus which interact with each other. In 3D space the wave function contains 107 degrees of freedom (DOF) ($26 * 3$ (electron coordinates) + $1 * 3$ (nucleus coordinates) + $(26 * 2)$ (if electron spin)). A further reduction can be achieved by the separation of the electrons in quasi-free and quasi-bound electrons. Quasi-bound electrons are located in the vicinity of the nucleus so that they are not able to “escape” out of the potential of the core. They do not affect the material properties (for example the electronic conductivity) like quasi-free electrons do. Coming back to the example of Fe, the subdivision into 18 quasi-bound and eight quasi-free electrons reduces the DOF to $8*3+1*3+1*8=35$.

Considering alloys with more than one species of atoms, the DOF are increasing drastically and furthermore for the description of point defects (≈ 10 to 100 atoms are needed). As is apparent from the consideration above, further simplifications are necessary. These simplifications are discussed below.

3.1 From the many-body problem to Density Functional Theory

The first important approximation (reduction of the problem) is the so called Born-Oppenheimer Approximation (BOA), which was first derived for a molecule in quantum chemistry and is named after Max Born and J. Robert Oppenheimer [12].

The Born-Oppenheimer Approximation

The time independent Schrödinger equation 3.1 of a molecule containing N_{nuc} nuclei and N_{el} electrons is a second order partial differential equation with $(N_{nuc} + N_{el}) \times 3 + (N_{el} \text{ if spin-decomposed})$

DOF.

The basic idea of the BOA is to reduce the dimension of the equation by separating the electron motion ψ_{el} from the nucleus motion ψ_{nuc} , thus the motion of electrons is considered within a static external nucleus potential with $\tilde{V}_{\text{el-nuc}} = \tilde{V}_{\text{ext}}$, and to solve the Schrödinger equation in two separate steps. The new wave function has then the form

$$\psi_{\text{BOA}}(\mathbf{r}, \mathbf{R}) = \psi_{\text{el}}(\mathbf{r}; \mathbf{R}) \cdot \psi_{\text{nuc}}(\mathbf{R}) \quad (3.7)$$

with \mathbf{R} , \mathbf{r} as the coordinates of the nuclei and electrons. The interaction between the nuclei and electrons $\tilde{V}_{\text{el-nuc}}$ can be assumed as an external potential, thus

$$\tilde{V}_{\text{el-nuc}}(\mathbf{r}; \mathbf{R}) =: \tilde{V}_{\text{ext}}(\mathbf{r}; \mathbf{R}) \quad (3.8)$$

and the electronic Hamilton operator is then in the form

$$\tilde{H}_{\text{el}}(\mathbf{r}; \mathbf{R}) = \tilde{T}_{\text{el}}(\mathbf{r}) + \tilde{V}_{\text{el-el}}(\mathbf{r}) + \tilde{V}_{\text{ext}}(\mathbf{r}; \mathbf{R}). \quad (3.9)$$

The electronic ground state can be derived from the equation

$$\tilde{H}_{\text{el}}(\mathbf{r}; \mathbf{R})\psi_{\text{el}}(\mathbf{r}; \mathbf{R}) = E_{\text{el}}(\mathbf{r}; \mathbf{R})\psi_{\text{el}}(\mathbf{r}; \mathbf{R}). \quad (3.10)$$

The wave function itself is often described by a linear combination of a orthonormal basis set

$$\psi_{\text{el}}(\mathbf{r}; \mathbf{R}) = \sum_i^n c_i \phi_i(\mathbf{r}) \quad (3.11)$$

with

$$\langle \phi_i | \phi_j \rangle = \delta_{ij}. \quad (3.12)$$

The energy of the electronic ground state $\psi_{\text{el},0}$ is then given by

$$E_{\text{el},0} = \langle \psi_{\text{el},0} | \tilde{H}_{\text{el}} | \psi_{\text{el},0} \rangle. \quad (3.13)$$

The Born-Oppenheimer energy surface is defined as

$$E_{\text{BOA}}(\mathbf{r}, \mathbf{R}) := E_{\text{el},0}(\mathbf{r}; \mathbf{R}) + V_{\text{nuc-nuc}}(\mathbf{R}), \quad (3.14)$$

and consequently the total energy of the whole system is then given by

$$E_{\text{tot}} = T_{\text{nuc}} + E_{\text{BOA}}. \quad (3.15)$$

The separation of the electronic motion and the ionic motion reduced the size of the many-body problem but due to the large number of electrons (especially their coordinates and spins) is the calculation of the Schrödinger equation 3.10 and hence the energy in equation 3.13 still impossible.

Further approximations for the electronic motion given by the so called Density-Functional Theory (DFT). The complex energy equation 3.13 is replaced by a calculation method which is

using the electronic density,

$$E_{\text{el},0} = \langle \psi_{\text{el},0} | \tilde{H}_{\text{el}} | \psi_{\text{el},0} \rangle = E[n] \quad (3.16)$$

i.e. an energy functional on the density n .

In 1964 Pierre Hohenberg and Walter Kohn proved that DFT is an exact theory of describing a many-body systems [13]. One year later in 1965 W. Kohn and Lu Jeu Sham introduced single-particle equations to solve the originally many-body system [97]. This will be presented in the following part.

Hohenberg-Kohn theorems

The precondition for the Hohenberg-Kohn theorems is the existence of an external potential $v_{\text{ext}}(\mathbf{r})$ which is not affected by the motion of the electrons.

“It is proved that there exists a universal functional of the density $n(\mathbf{r})$, independent of the external potential $v(\mathbf{r})$, such that the expression

$$E \equiv \int v_{\text{ext}}(\mathbf{r})n(\mathbf{r})d\mathbf{r} + F[n(\mathbf{r})] \quad (3.17)$$

has as its minimum value the correct ground-state energy associated with $v_{\text{ext}}(\mathbf{r})$ ” [13].

Here E denotes the electronic ground state energy E_{el} mentioned in equation 3.13, and the right-hand side of the equation denotes the Hohenberg-Kohn energy functional $E_{\text{HK}}[n(\mathbf{r})]$ on the charge density $n(\mathbf{r})$. E_{el} is equal to $E_{\text{HK}}[n]$ when the charge density $n(\mathbf{r})$ corresponds to the true unknown charge density distribution (without excitation). If this is the case, the integral in the equation 3.17 corresponds to the Coulomb energy of electron-nuclei interaction. The second term, the density functional $F[n(\mathbf{r})]$ will be explained below.

“The external potential $v_{\text{ext}}(\mathbf{r})$ for any system is determined, except of a constant shift, uniquely by the electronic ground state density $n_0(\mathbf{r})$ ”.

In other words, the electronic density n determines the external potential $v_{\text{ext}}(\mathbf{r})$ and all ground state properties. This is schematically shown in figure 3.2.

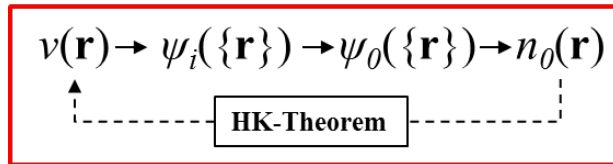


Figure 3.2: Schematic drawing of the Hohenberg-Kohn circle. The external potential $v_{\text{ext}}(\mathbf{r})$ is determined, except for a constant, by the ground state density $n_0(\mathbf{r})$.

The theorem is a prove that a energy functional on the density $E_{\text{HK}}[n]$ (which gives the ground-state energy E_0 by minimizing it) must exist, but doesn't offer how to obtain the ground-state energy. It writes

$$E_0 \leq E[n_i] \quad (3.18)$$

and

$$E_0 = E[n_0] = \min_n E_{\text{HK}}[n]. \quad (3.19)$$

Similar to the terms of the sum of equation 3.9 the Hohenberg-Kohn energy functional $E_{\text{HK}}[n]$ contains the external energy functional $E_{\text{ext}}[n]$, the electron kinetic energy functional $T[n]$, and the repulsive electron-electron interaction functional $V_{\text{int}}[n]$ on the electron density. Considering equation 3.17 and defining

$$E_{\text{ext}}[n] = \int v_{\text{ext}}(\mathbf{r})n(\mathbf{r})d\mathbf{r} \quad (3.20)$$

it becomes clear that $F_{\text{HK}}[n]$ can be expressed as the sum

$$F_{\text{HK}}[n] = T[n] + V_{\text{int}}[n]. \quad (3.21)$$

The potential functional in equation 3.21 contains a classical term, the so called Hartree term $E_H[n]$, and a non-classical term $E_{xc}[n]$. The Hartree term describes the classical repulsive (long range) Coulomb energy of the electron-density which is interacting with itself

$$E_H[n] = \frac{1}{2} \int \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}d\mathbf{r}'. \quad (3.22)$$

The non-classical term, namely the energy functional density of the exchange and correlation of the electrons, is yet unknown and must be approximated. This will be discussed in more detail below. Equation 3.21 can then be reformulated as

$$F_{\text{HK}}[n] = T[n] + E_H[n] + E_{xc}[n], \quad (3.23)$$

and hence the Hohenberg-Kohn energy functional in equation 3.19

$$E_{\text{HK}}[n] = T[n] + E_{\text{ext}}[n] + E_H[n] + E_{xc}[n]. \quad (3.24)$$

Kohn-Sham Equations

In the Kohn-Sham equations the many-body system of N_{el} interacting electrons is replaced by an auxiliary system of non-interacting electrons, assuming that the ground state density of interacting electrons is equal to the ground state density of specific chosen system of non-interacting electrons in an effective potential.

This simplification allows to solve N_{el} single-particle Schrödinger equations in the form

$$\tilde{H}_{\text{KS}} |\varphi_i\rangle = \left(-\frac{1}{2} \nabla_i^2 + v_{\text{eff}}(\mathbf{r}) \right) |\varphi_i\rangle = \epsilon_i |\varphi_i\rangle \quad (3.25)$$

under the constraint of orthogonal single-particle wave functions

$$\langle \varphi_i | \varphi_j \rangle = \delta_{ij} \quad (3.26)$$

in an effective potential $v_{\text{eff}}(\mathbf{r})$. According to equation 3.24 the effective potential writes

$$v_{\text{eff}}(\mathbf{r}) = \frac{\delta E_{\text{ext}}}{\delta n(\mathbf{r})} + \frac{\delta E_H}{\delta n(\mathbf{r})} + \frac{\delta E_{xc}}{\delta n(\mathbf{r})} = v_{\text{ext}}(\mathbf{r}) + v_H(\mathbf{r}) + v_{xc}(\mathbf{r}), \quad (3.27)$$

and the density

$$n(\mathbf{r}) = \sum_i^{N_{el}} \sum_s |\varphi_i(\mathbf{r}, s)|^2 \quad (3.28)$$

in which s denotes the spin. The Kohn-Sham equations 3.25, 3.27, and 3.28 depending on each other and must be solved self-consistently. This is schematically shown in figure 3.3

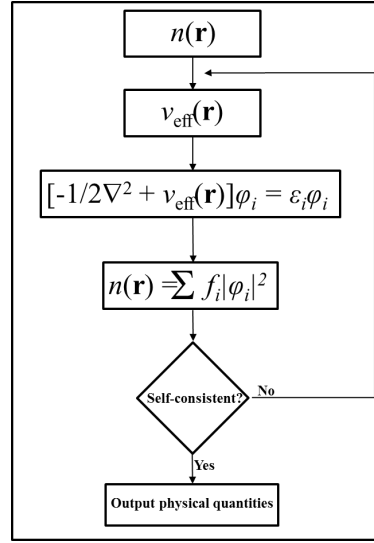


Figure 3.3: Schematic drawing of the self-consistency loop of the Kohn-Sham equations. Starting with an initial guess for the density, the effective potential $v_{\text{eff}}(\mathbf{r})$ is determined through it. N_{el} single-particle Schrödinger equations with the given effective potential are solved self-consistently.

In equation 3.27 all terms of the sum, except the exchange-correlation potential $v_{xc}(\mathbf{r})$ are well defined. The latter term needs to be approximated. This will be discussed in the following part.

Exchange-correlation Functional (GGA and LDA)

In the mentioned Kohn-Sham loop the only unknown part is the exchange-correlation potential which is the derivative of the exchange-correlation functional with respect to the density n

$$v_{xc}(\mathbf{r}) = \frac{\delta E_{xc}[n]}{\delta n(\mathbf{r})}. \quad (3.29)$$

The exact functional is the difference in the kinetic and potential energy between the true interacting many-body system and the Hartree and kinetic energy of the auxiliary system

$$E_{xc}[n] = \langle \tilde{T} \rangle - T[n] + \langle \tilde{V} \rangle - E_H[n]. \quad (3.30)$$

Since the true energies are unknown, this functional must be approximated.

Local Density Approximation

Historically the first approximation for this functional was performed by Hohenberg and Kohn with the so called Local Density Approximation (LDA) assuming that exchange and correlation of electrons are local effects and that the energy density is the same as in a homogeneous electron gas ϵ_{hom} . It writes

$$E_{xc}^{\text{LDA}}[n] = \int n(\mathbf{r}) \epsilon_{\text{hom}}(n(\mathbf{r})) d\mathbf{r}, \quad (3.31)$$

and its functional derivative

$$\frac{\delta E_{xc}[n]}{\delta n(\mathbf{r})} = v_{xc}(\mathbf{r}) = \epsilon_{\text{hom}}(\mathbf{r}) + n(\mathbf{r}) \frac{\partial \epsilon_{\text{hom}}(n)}{\partial n(\mathbf{r})}. \quad (3.32)$$

LDA is the simplest approximation for the exchange-correlation term. An extension of LDA is the so called Generalized-Gradient Approximation (GGA) which is using the gradient of the density $|\nabla n|$ as further information. This extension leads in general to an improvement for most but not for all material systems.

Generalized-Gradient Approximation

All developed GGA functionals (PBE [98], PW91 [99], B88 [100], ...) can be expressed in the general form

$$E_{xc}^{\text{GGA}}[n] = \int n(\mathbf{r}) \epsilon_{\text{hom}}(n, |\nabla n|) d\mathbf{r}, \quad (3.33)$$

and the potential in the general form

$$v_{xc}(\mathbf{r}) = \epsilon_{\text{hom}}(\mathbf{r}) + n(\mathbf{r}) \frac{\partial \epsilon_{\text{hom}}(n)}{\partial n(\mathbf{r})} - \nabla \left(n \frac{\partial \epsilon_{\text{hom}}(n)}{\partial \nabla n} \right). \quad (3.34)$$

The GGA functional used in this work is PBE, developed by Perdew, Burke, and Ernzerhof in 1996.

The simplifications mentioned above reduce the problem of solving the originally formulated Schrödinger equation 3.1 but still do not solve the huge number of interacting electrons. As already mentioned the periodicity of solids turns out to be a huge advantage to reduce the problem drastically by using Bloch-functions. This will be explained in the following part.

Periodic Boundary Condition and Bloch's theorem

The wave function of an electron in a potential with \mathbf{R} -periodic boundary conditions (i.e. $v(\mathbf{r}) = v(\mathbf{r} + \mathbf{R})$) can be expressed as a product of a plane wave $\exp(i\mathbf{k}\mathbf{r})$ and a \mathbf{R} -periodic function $u_{\alpha\mathbf{k}}(\mathbf{r}) = u_{\alpha\mathbf{k}}(\mathbf{r} + \mathbf{R})$

$$\psi_{\alpha\mathbf{k}}(\mathbf{r}) = u_{\alpha\mathbf{k}}(\mathbf{r}) \exp(i\mathbf{k} \cdot \mathbf{r}) = u_{\alpha\mathbf{k}}(\mathbf{r}) (\cos(\mathbf{k} \cdot \mathbf{r}) + i \sin(\mathbf{k} \cdot \mathbf{r})), \quad (3.35)$$

with $\mathbf{R} = n_1 \mathbf{a}_1 + n_2 \mathbf{a}_2 + n_3 \mathbf{a}_3$ where \mathbf{a}_i denotes the three basis vectors of a Bravais lattice and n_i integer numbers.

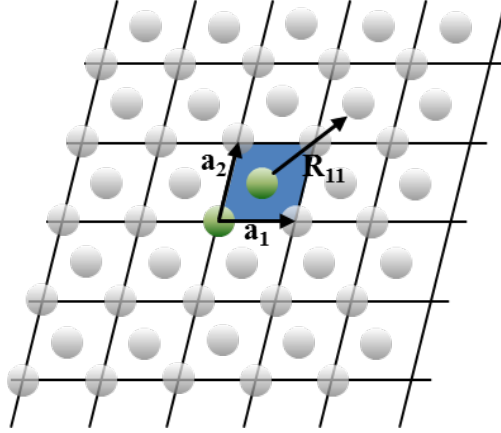


Figure 3.4: 2D example of a periodic lattice with two basis atoms in the unit cell (blue shaded). \mathbf{a}_1 and \mathbf{a}_2 denotes the two Bravais vectors of the cell. The translation vector $\mathbf{R}_{11} = 1\mathbf{a}_1 + 1\mathbf{a}_2$ is one example out of an infinite number of possible translations. The wave functions of the electrons in such a crystal can be expressed as Bloch functions.

Here \mathbf{k} denotes the wave vector which is defined in the first Brillouin zone and α counts the eigenvalues to the corresponding vector \mathbf{k} . Equation 3.35 is also known as so called Bloch-function (named after Felix Bloch) [101]. The periodic function $u_{\alpha\mathbf{k}}(\mathbf{r})$ can be expressed through a Fourier transformation in the reciprocal space

$$u_{\alpha\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{G}} c_{\alpha\mathbf{k}}(\mathbf{G}) \exp(i\mathbf{G} \cdot \mathbf{r}), \quad (3.36)$$

with \mathbf{G} being a reciprocal lattice vector. Considering both equations 3.35 and 3.36 the single-particle wave function of an electron in a periodic potential can be expressed as

$$\psi_{\alpha\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{G}} c_{\alpha\mathbf{k}}(\mathbf{G}) \exp(i(\mathbf{k} + \mathbf{G}) \cdot \mathbf{r}). \quad (3.37)$$

Bloch's theorem allows to reduce the description of the electronic wave function of a quasi-infinite crystal to the limited region of a unit cell (Brillouin zone).

The wave functions of the valence electrons are drastically fluctuating close to the core due to the requirement of being orthogonal to the core electrons (number of plane waves in a basis-set are increasing to describe the fluctuation accurately). In this manner solving single-particle wave functions under the constraint of orthogonality becomes in an all-electron approach computationally demanding especially for the valence electrons.

The last simplification is performed with the separation of quasi-free and quasi-bound electrons. The so called pseudopotential approximation (in contrast to an all-electron approach) was introduced by Hans Hellmann 1935 [102] and simplifies the solution of the electronic Schrödinger equations drastically (number of plane waves in a basis-set are decreasing).

Pseudopotentials

Because of screening effects and a very strong electron-nucleus Coulomb interactions electrons close to the core are “trapped” and do not influence material properties very much (they are chemically inactive). The wave function close to the core region, which is defined by a cutoff-radius, is approximated by smooth functions in order to flatten out the wave function in this region. The wave function above the cutoff radius of the approximated system is equal to the wave function of the all-electron system (see figure 3.5). The mentioned approximation - the separation of core and valence electrons and the smoothing of the wave functions for the valence electrons in the core region - leads to the last simplification of the originally formulated many-body problem.

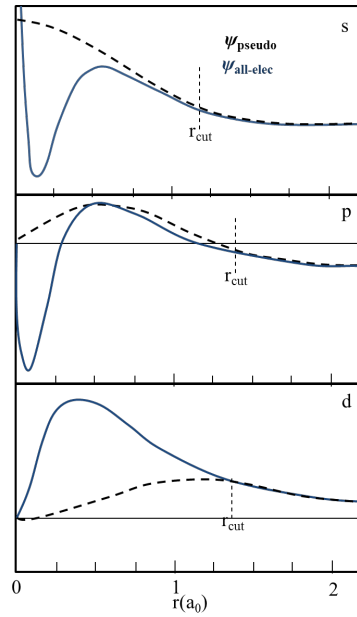


Figure 3.5: s, p, d pseudo-wave functions (dashed line) and all-electron wave functions (blue line) of Mn obtained in the valence band region (-8.16 eV). Above the cutoff radius r_{cut} the wave functions of the pseudo system corresponds to the all-electron system [103].

3.2 Solubility of interstitial elements

A central motivation of the HT study in this thesis is the fact that interstitial elements can significantly influence the materials properties of a host material. One of the central questions to approach these phenomena, is the number of interstitial atoms that can be dissolved in the bulk material or in other words the solution enthalpy of interstitial atoms. In this context, hydrogen serves as a prototype interstitial element for our investigations.

The chemical reaction of originally $\frac{n}{2}$ diatomic gas molecules X_2 (e.g., hydrogen molecules) dissolved as n single atoms in a bulk system (e.g., a transition metal) in the form



is exothermic, if the system releases heat ($\delta Q > 0$) and endothermic if the system requires heat for the transformation ($\delta Q < 0$). The product system is for an exothermic reaction energetically more favorable than the reactant system (see figure 3.6).

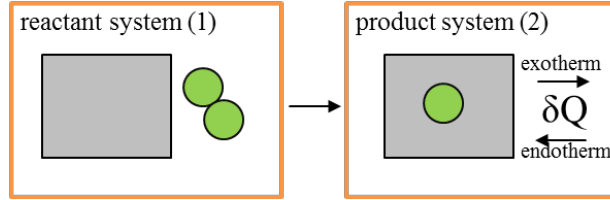


Figure 3.6: The reactant system contains the bulk contribution (a transition metal) and a single X_2 molecule. The product system shows the dissolved single atom in the bulk. If the product system is more favorable, the heat $\delta Q > 0$ is released. With δQ the interstitial solubility is increasing.

For constant temperature $T = T_0$, constant pressure $p = p_0$, and a fixed number of interstitial atoms, the heat is equal to the change of the enthalpy $\delta Q = -\Delta H$. Assuming a very low interstitial concentration (ppm range) that is dissolved in the metal, implying that the individual interstitial atoms do not interact with each other the already introduced equation 1.1 of the form

$$c_H = \sqrt{\frac{p}{p_0}} \cdot e^{\frac{\Delta S}{k_B}} \cdot e^{-\frac{\Delta H}{k_B T}} \quad (3.39)$$

can be derived [55]. Here T is the temperature, p is the pressure, p_0 the reference pressure (in general $p_0 = 1$ atm), k_B the Boltzmann constant, ΔS is the entropy change between the reactant and product system, which is assumed to be constant with respect to the temperature T . One should note that this assumption is not necessarily fulfilled in reality. Various point defects show deviations from a temperature independent entropy of formation for point defects. In the case of vacancies, for example, a recent study [104] has shown a strongly nonlinear temperature dependence of the free energy of formation, resulting from anharmonic lattice vibrations. However, at room temperature anharmonic effects are negligible small and therefore within a narrow temperature regime the assumption of a constant entropy will be justified. The key quantity in equation 3.39 is the already mentioned solution enthalpy ΔH , which is the enthalpy difference between the product system (2) and the reactant system (1) as shown in figure 3.6

$$\Delta H_{1 \rightarrow 2} = H_2 - H_1. \quad (3.40)$$

The enthalpy is the sum of the internal energy U and the product of the pressure p and the volume V

$$H = U + pV. \quad (3.41)$$

The solution enthalpy can be determined experimentally as follows: If the interstitial containing metallic phase is in thermodynamic equilibrium with a surrounding atmosphere of gaseous interstitial molecules, then the chemical potential of a single interstitial atom $\frac{1}{n}\mu_X^g$ in gas atmosphere needs

to be equal to the chemical potential of the atomic interstitial dissolved in the host metal μ_X^M , i.e.

$$\frac{1}{m}\mu_X^g = \mu_X^M, \quad (3.42)$$

in which m denotes the number of atoms per molecule (for example $m = 2$ in the case of a diatomic hydrogen molecule). From the equilibrium condition the Van 't Hoff equation for the solution enthalpy

$$\Delta H_{1 \rightarrow 2} = \frac{R}{2} \frac{\partial \ln p}{\partial (1/T)} \quad (3.43)$$

can be derived in dependence of the ideal gas constant R , the dissociation pressure p , and the temperature T (see figure 3.7). Each dot in the figure corresponds to several isotherm experimental measurements. In contrast to the sophisticated experimental method, the determination of ΔH with *ab initio* methods is straight forward and will be presented in the following section.

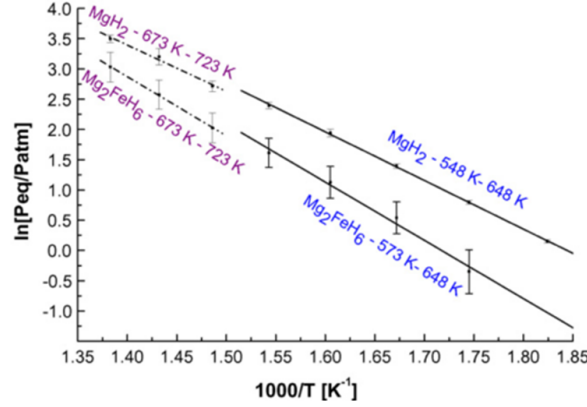


Figure 3.7: Copied from [105]. Van 't Hoff plot of the complex hydride Mg_2FeH_6 and MgH_2 . Black dots belong to pressure composition isotherms measured at different temperatures. According to equation 3.43 the slopes corresponds to ΔH .

3.3 *Ab initio* description of the solution enthalpy

The calculation of the ΔH within DFT is conceptionally straight forward. It is the energy difference between the relaxed interstitial-metal system and the reference pure metallic system and the chemical potential of the interstitial element. Considering equations 3.40 and 3.41 the solution enthalpy of the metal-interstitial system after interstitial solution at zero pressure is given as

$$\Delta H_0 = E_{MX} - \left(E_M + \frac{1}{2}\mu_{X_2} \right), \quad (3.44)$$

in which E_{MX} denotes the relaxed energy of the transition-metal super-cell containing a single interstitial atom, E_M denotes the reference energy of the pure transition-metal super-cell, and $\frac{1}{2}\mu_{X_2}$ corresponds to the half energy of a diatomic interstitial molecule in vacuum. In order to determine

the energies of the distorted and pure system both must be calculated at zero pressure $p = 0$. In order to obtain the energies at corresponding zero pressures two methods are implemented and discussed in the results chapter 5. The chemical potential of the interstitial elements are also calculated at $p = 0$. This is simply a X_2 molecule in vacuum. However, one should note that the chemical potential of the interstitial element can vary according to its environment (e.g. temperature or pressure) [106]. The current thesis is focused on interstitial elements in dilute limit. For this purpose, the size of the supercell must be checked carefully to avoid interstitial-interstitial interaction with mirror-cells.

3.4 Zero point energy: Einstein method and dynamical matrix

For small interstitial atoms such as hydrogen, but also for the host crystal structure quantum effects due to zero-point vibrations might be relevant. They result into an energy shift of the ground state energy as compared to a classical treatment of the crystal structure. The zero-point energy (ZPE) E_{ZPE} must in principle be added to each term in equation 3.44. For $\Delta E_{\text{ZPE}} = E_{\text{MX}}^{\text{ZPE}} - (E_{\text{M}}^{\text{ZPE}} + \frac{1}{2}\mu_{X_2}^{\text{ZPE}})$ equation 3.44 can then be reformulated as

$$\Delta H_{0+\text{ZPE}} = \Delta H_0 + \Delta E_{\text{ZPE}}. \quad (3.45)$$

The most simple description of ZPE is achieved, if the lattice vibrations of a single particle are mapped on a one-dimensional harmonic oscillator, for which

$$E_{\text{ZPE}}^{\text{sp}} = \frac{1}{2}\hbar\omega, \quad (3.46)$$

and where \hbar is the reduced Planck's constant and ω denotes the oscillation frequency of the particle. More generally, to determine the energy for the (coupled) oscillations in a crystal exactly, the whole phonon dispersion $\omega_{\mathbf{q}}$ within the first Brillouin zone must be known. The ZPE for the crystal is then defined as

$$E_{\text{ZPE}}^{\text{cell}} = \frac{1}{2}\hbar \int^{\Omega_{\text{BZ}}} \omega_{\mathbf{q}} d\mathbf{q}. \quad (3.47)$$

In this work the ZPE is determined in two different ways, within the Einstein approach [107] and by calculating the full dynamical matrix.

Einstein method

The Einstein method is computationally less demanding, since it describes the ZPE of the crystal by a single effective frequency ω . Here, we assume that the difference between the ZPE of the metal-interstitial system and the ZPE of the pure host material is dominated by the vibrations of the single interstitial atom, which is missing in the pure host material. Therefore, it is sufficient to determine the vibrational frequencies of the interstitial atom in the harmonic approximation using the potential

$$V(x, y, z) = \frac{1}{2}m (\omega_x^2 x^2 + \omega_y^2 y^2 + \omega_z^2 z^2), \quad (3.48)$$

in which ω_{α} is the vibrational frequency of the interstitial atom in $\alpha = x, y, z$ direction, respectively.

The ab initio energy is determined for small displacements of the interstitial atom out of its

equilibrium position into three perpendicular directions and a second-order polynomial fit is used to determine the vibrational frequencies

$$\omega_{x,y,z} = \sqrt{\frac{2a_{x,y,z}}{m}}. \quad (3.49)$$

with a_α as quadratic pre-factor and the total frequency

$$\omega_{MX} = \omega_x + \omega_y + \omega_z \quad (3.50)$$

of the interstitial in the host material. As already mentioned above the zero-point energy difference between the metal-interstitial system and the pure host material can be approximated with this frequency, thus

$$E_{ZPE}^{MX} - E_{ZPE}^M = \frac{1}{2}\hbar\omega_{MX}. \quad (3.51)$$

The last term of equation 3.45 can then be completed with the ZPE of the X_2 molecule

$$\Delta E_{ZPE} = E_{ZPE}^{MX} - E_{ZPE}^M - \frac{1}{2}E_{ZPE}^{X_2} = \frac{1}{2}\hbar\left(\omega_{MX} - \frac{1}{2}\omega_{X_2}\right) \quad (3.52)$$

in which $E_{ZPE}^{X_2}$ correspond to the zero-point energy of the X_2 molecule and ω_{X_2} denotes the vibration frequency of it.

Dynamical matrix

The computationally more demanding approach is the calculation of the full dynamical matrix of the supercell. In contrast to the Einstein method, this requires the displacement of all non-equivalent atoms in the supercell and the determination of the force components on the atoms that result from the displacement. In the harmonic approximation a proportionality $F_i = -\sum_j \Phi_{i,j}u_j$ of the forces F_i on an atom i after the displacement of atoms j is assumed. Here, $\Phi_{i,j}$ is the force constant matrix and the indices distinguish also special directions. Analogous to equation 3.49, the diagonalization of the matrix allows the determination of the eigen frequencies of the atoms

$$\omega_i = \sqrt{\frac{\Phi_{ii}^{\text{diag}}}{m_{ii}}}, \quad (3.53)$$

with $i = 1, \dots, 3N$ and Φ_{ii}^{diag} denotes the matrix elements of the diagonalized matrix. The diagonalization yields a coupling of all oscillators in terms of phonon modes. Therefore, the displacement of a single atom affects all other atoms in the supercell. The displacement of all atoms leads to the general formulation of the force constant matrix

$$\Phi_{i,j} := -\frac{\partial F_{\mathbf{u}_j}^i}{\partial \mathbf{u}_j} \approx -\frac{F_{\mathbf{u}_j}^i}{|\mathbf{u}_j|}, \quad (3.54)$$

with $i = 1, \dots, 3N$ and $j = 1, \dots, 3N$. [108] [109]. After Fourier transformation of Φ , which is called the dynamical matrix, the periodic boundary conditions of the supercell are taken into account.

If the supercell (containing N atoms) consists of P primitive cells, then the originally formulated $(3N \times 3N)$ -dimensional force constant matrix can be replaced by $(3r \times 3r)$ -dimensional (with $r = N/P$) matrices $\hat{\Phi}^P$ of P primitive cells and their positions \mathbf{R}_P (i.e. the translation vector w.r.t. the first primitive cell defined at the origin). The dispersion relation $\omega_{\mathbf{q}}$ can be determined by diagonalization of the dynamical matrix

$$D_{\alpha,\beta}(\mathbf{q}) = \frac{1}{\sqrt{m_{\alpha}m_{\beta}}} \sum_P \hat{\Phi}_{\alpha,\beta}^P \exp(i\mathbf{q} \cdot \mathbf{R}_P), \quad (3.55)$$

with the masses of the atoms m_{α}, m_{β} , the force constant matrices $\hat{\Phi}^P$ and their positions \mathbf{R}_P of the primitive cells.

Due to symmetry, this dispersion relation of the phonon frequencies $\omega_{\mathbf{q}}$ can be limited to the first Brillouin zone of the primitive cell (containing r atoms) only. The number of accurate frequencies (at exact \mathbf{q} points) is increasing with increasing the numbers of primitive cells in the supercell. Since we consider a dilute concentration of interstitial atoms in our work, the translational symmetry of the supercell is broken. The required large supercells will contain only one exact \mathbf{q} point. At the same time they are connected with a very small Brillouin zone. Therefore, the determination of the frequencies at the Γ point is considered as being sufficiently accurate for the present purpose.

The vibrational contribution to the solution enthalpy of interstitial atoms is typically ignored in the literature. For the interstitial element hydrogen, however, quantum effects like ZPE corrections can be relevant for the diffusivity [40, 110, 111] and also for the solubility [112, 113], since it is a light element and experiences a rather stiff potential in the interstitial position of a transition metal. Therefore, they might also affect the solubility significantly. Similar to the Einstein approach 3.51, the ZPE contribution of each phonon mode is now added to obtain

$$E_{\text{ZPE}}^{\text{MX}} = \frac{1}{2} \hbar \sum_{i=1}^{3N+3} \omega_{ii}^{\text{Diag}}(\mathbf{q} = (0, 0, 0)) \quad (3.56)$$

for the metal-hydrogen system and

$$E_{\text{ZPE}}^{\text{M}} = \frac{1}{2} \hbar \sum_{i=1}^{3N} \omega_{ii}^{\text{Diag}}(\mathbf{q} = (0, 0, 0)) \quad (3.57)$$

for the pure metallic system. The total ZPE correction can again be calculated according to equation 3.52.

3.5 Semi-empirical models for interstitial solubilities

The general philosophy of an approach to predict interstitial solubilities in transition metals is explained in this section. For this purpose, we choose hydrogen as a prototype interstitial element to understand the mechanisms governing the solubility. Furthermore, we describe the only semi-empirical model we are aware of that predicts the hydrogen solubility based on a systematic study of experimental data. It has been introduced by Griessen *et al.* [114].

The desired model is assumed to describe the target property "solution enthalpy" as a function

$f(X)$, i.e.

$$\Delta H = f(X), \quad (3.58)$$

of a yet unknown combinations of parameters X that contains information about the host material and/or the interstitial element, but does not use any information about the interstitial in the host as input.

Griessen's model is special in the sense that it has with the "volume expansion" of the host material due to hydrogen incorporation another target property. It is best described by the partial molar volume \bar{V}_H which is by definition the change of host volume with respect to the change of moles hydrogen $\partial V / \partial n_H$ in the host [115]. Considering atomic units and according to subsection 3.3, the excess volume by definition is given as

$$\bar{V}_H = \frac{V_{MX} - V_M}{n_X}, \quad (3.59)$$

in which V_{MX} denotes the equilibrium volume of the metal-interstitial system, V_M denotes the equilibrium volume of the pure host system and n_X denotes the number of interstitial atoms in the host (which is in practical calculations equal to one, if the dilute limit is considered).

Griessen and Feenstra [115] demonstrated that in the dilute limit the change of the solution enthalpy of hydrogen with respect to the logarithmic change of the host materials volume for constant temperature is very well described by the approximation

$$\frac{\partial \Delta H}{\partial \ln V} \approx -B\bar{V}_H, \quad (3.60)$$

i.e. by the product of the excess volume of the hydrogen atom and the bulk modulus of the host material. Equation 3.60 is for the dilute limit considered here actually almost exact [115]. The correlation of the such predicted with experimental values for the solution enthalpy is shown in figure 3.8. It has been determined for a large variety of host materials and very low hydrogen concentrations.

Taking both equations 3.58 and 3.60 into account, it follows

$$B\bar{V}_H = \frac{df}{d \ln X} \left(-\frac{\partial \ln X}{\partial \ln V} \right) =: \frac{df}{d \ln X} \gamma_X. \quad (3.61)$$

If γ_X is constant for all unknown parameters X it directly follows

$$B\bar{V}_H = \varphi(\Delta H) \quad \text{or} \quad \frac{df}{d \ln X} \gamma_X = \varphi(f(X)). \quad (3.62)$$

The partial differential equation 3.62 can be rewritten as

$$\frac{df}{\varphi(f(X))} = \frac{1}{\gamma_X} d \ln X. \quad (3.63)$$

Since $d \ln X$ is equal to $\frac{1}{X} dX$, the integration of both sides yields

$$\int_{f_0}^f \frac{df}{\varphi(f)} = \int_{X_0}^X \frac{1}{\gamma_X} \frac{1}{X} dX \quad (3.64)$$

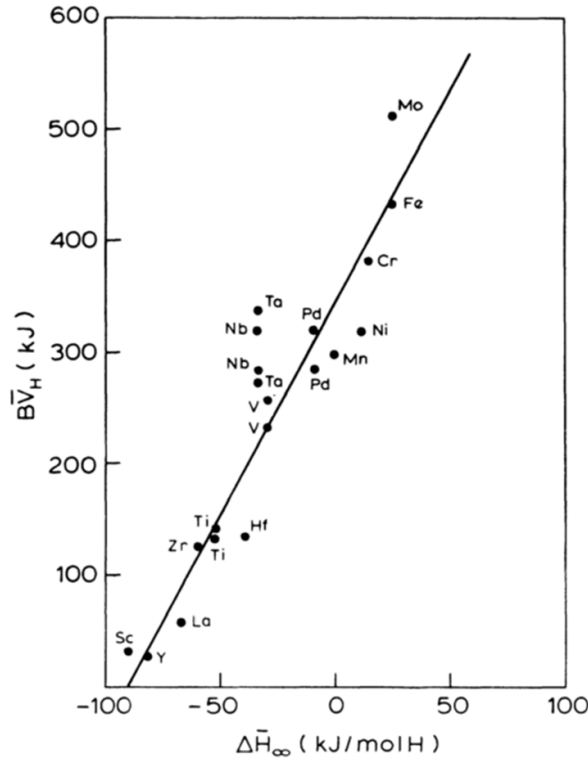


Figure 3.8: Copied from [114]. Comparison between the experimental obtained solution enthalpy of hydrogen in 3d and 4d transition metals and the product of the bulk modulus and the excess volume of hydrogen in the reference host metal.

Having in mind that we considered γ_X to be constant, the right-hand side of equation 3.64 can be solved by using analytic properties of logarithms:

$$\int_{f_0}^f \frac{df}{\varphi(f)} = \ln \left(\frac{X}{X_0} \right)^{\frac{1}{\gamma_X}}. \quad (3.65)$$

We now determine the function $\varphi(f)$ from a fit through the experimental data shown in figure 3.8. From this we get

$$\varphi(f) = 3.83f + 345. \quad (3.66)$$

Next we transform the left-hand side of equation 3.65

$$\int_{f_0}^f \frac{1}{3.83f + 345} df = \frac{1}{3.83} \ln \left(\frac{3.83f + 345}{3.83f_0 + 345} \right) = \ln \left(\frac{3.83f + 345}{3.83f_0 + 345} \right)^{\frac{1}{3.83}} = \ln \left(\frac{X}{X_0} \right)^{\frac{1}{\gamma_X}}. \quad (3.67)$$

By taking the exponential on both sides, it follows

$$\left[\frac{X}{X_0} \right]^{\frac{1}{\gamma_X}} = \left[\frac{3.83f + 345}{3.83f_0 + 345} \right]^{\frac{1}{3.83}} \quad (3.68)$$

and comparing both sides one can get the value $\gamma_X = 3.83$ and

$$X = k(3.83f + 345) \quad \text{or} \quad \eta X = f + 90 \quad (3.69)$$

with unknown constants k or $\eta = \frac{1}{3.83k}$. It was therefore Griessen's approach to find a product of physical parameters X of the pure host system which linearly correlates with the function f under the constraint $\gamma_X = 3.83$.

Earlier energy level analysis [116] on Pd_6H_1 clusters revealed that the incorporation of hydrogen reduces the lowest lying conduction band of the metal phase (highlighted at bottom of figure 3.9), but the energy reduction is almost compensated by the repulsive electron Coulomb interaction due to an increased charge density close to the protons. Overall the system energy is decreased and

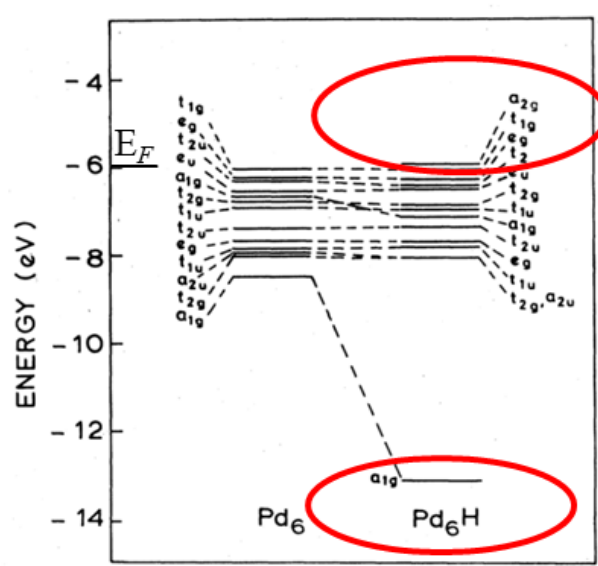


Figure 3.9: Copied from [116]. Band analysis of an octahedral Pd cluster achieved with a self-consistent field method (SCF-X α -SW). The insertion of hydrogen interstitial leads to a lowering of the lowest lying band of the host element. The electron coming from the hydrogen atom accommodates just above the Fermi energy of the Pd_6 cluster (red ellipses).

the hydride formation becomes beneficial. On the other hand, the additional electron coming from hydrogen is accommodated just above the Fermi level of the host material (highlighted at top of figure 3.9). The electron uptake at the Fermi level increases the system energy and hydride formation becomes expensive. The energy gain of the system is supposed to be the energy difference of the Fermi level and the energy of the center of the lowest lying conduction band of the pure Pd_6 cluster $\Delta E = E_F - E_s$. To achieve a model that takes the impact of local effects, such as volume expansion or the site occupancy, into account, Griessen introduced two more parameters, the d-band width W_d of the host element, which correlates with the d-d band overlap integral, and the distance of the interstitial site to the neighboring atoms R_j , i.e. the spacing of the interstitial site. Both affect

the solution enthalpy. The mentioned parameter X is then in the form

$$X = \Delta E \times W_d^m \times \sum_j R_j^{-n} \quad (3.70)$$

with the yet unknown exponentials n and m , and with j counting the nearest neighboring atoms (four in the case of tetrahedral site occupation and six in the case of octahedral site occupation).

Considering equation 3.61, γ_X must fulfill the constraint

$$\gamma_X = -\frac{\partial \ln X}{\partial \ln V} = -\frac{\partial \ln \Delta E_{F-s}}{\partial \ln V} - \frac{\partial \ln W_d^m}{\partial \ln V} - \frac{\partial \ln \sum_j R_j^{-n}}{\partial \ln V} = 3.83. \quad (3.71)$$

By using the correlations for $\Delta E_{F-s} \sim V^{-5/3}$, $W_d \sim V^{-5/3}$, and $R_j \sim V^{1/3}$, equation 3.71 can be reformulated as

$$3.83 = \frac{5}{3} + \frac{5}{3}m + \frac{1}{3}n, \quad (3.72)$$

or

$$6.5 = 5m + n \quad (3.73)$$

which is the constraint for solving the exponentials n and m of equation 3.70.

Satisfying equation 3.60 with the constraint $5m + n = 6.5$ and using a least squared errors fit to the experimental data shown in figure 3.8, the best result is achieved for $m = \frac{1}{2}$, $n = 4$ and the multiplicative constant $\eta = 18.6 \text{ (kJ/mol H)}(\text{\AA}^4\text{eV}^{-3/2})$. The predicting formula for the solution enthalpy is then given as

$$\Delta H = 18.6 \text{ (kJ/mol H)}(\text{\AA}^4\text{eV}^{-3/2})\Delta E W_d^{1/2} \sum_j R_j^{-4} - 90 \text{ kJ/mol H}. \quad (3.74)$$

The correlation between the predictive formula and the experimental values is presented in figure 3.10.

In the present work, it is assumed that such a predictive formula (which can also be a similar one with different fitting values) can be found not only for hydrogen, but also for all other interstitial elements under focus. Instead of using experimental input as shown in figure 3.8, we fully rely on ab initio calculations. Furthermore, we replace the analytic approach presented above, by an automated data analysis, which is based on concepts of multivariate-statistics.

3.6 Bravais-Pearson correlation coefficient

An important statistical tool for our analysis is the detection of linear correlations between different physical quantities, as for example the solution enthalpy of two different interstitial elements. For this purpose, the Bravais-Pearson correlation coefficient (PCC) is used, which is well established in statistical mathematics [117]. For two sets of data $\mathbf{a} = (a_1, a_2, \dots, a_n)$ and $\mathbf{b} = (b_1, b_2, \dots, b_n)$, which can be the solution enthalpies of two different interstitial elements (say a_i and b_i) for a series of

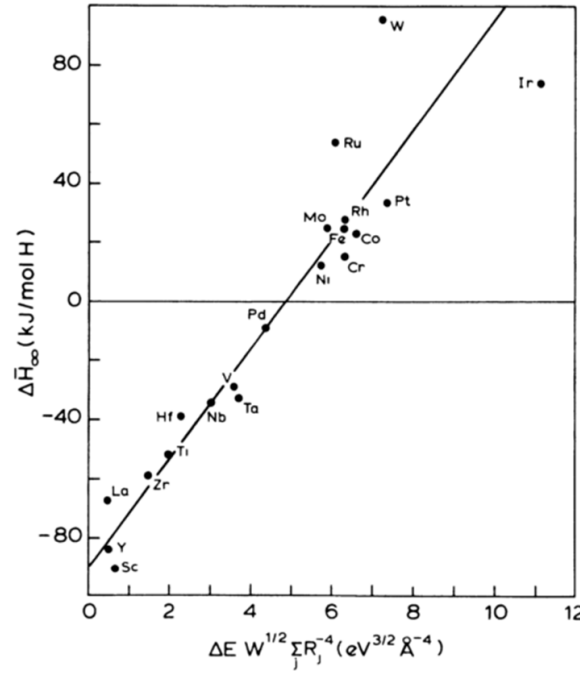


Figure 3.10: Copied from [114]. Griessen's predictive formula of the hydrogen solution enthalpy vs experimental values in 3d and 4d transition metals. An overall good accuracy can be observed.

host elements i , the PCC is defined as

$$r_{\mathbf{a},\mathbf{b}} = \frac{\text{cov}(\mathbf{a}, \mathbf{b})}{\sigma_{\mathbf{a}}\sigma_{\mathbf{b}}} = \frac{\sum_{i=1}^n (a_i - \langle \mathbf{a} \rangle) (b_i - \langle \mathbf{b} \rangle)}{\sqrt{\sum_{i=1}^n (a_i - \langle \mathbf{a} \rangle)^2} \sqrt{\sum_{i=1}^n (b_i - \langle \mathbf{b} \rangle)^2}}, \quad (3.75)$$

with the covariance of \mathbf{a} and \mathbf{b} in the nominator and their standard deviations in the denominator (having in mind that the coefficient $\frac{1}{n}$ in the nominator and denominator is canceled out). For a high correlation r is equal to one, for no correlation r will be close to zero, and for anti-correlation equal to minus 1. This is shown schematically in figure 3.11.

If more than two physical quantities are compared to each other it is more convenient to use the matrix notation and calculate the PCC matrix for all combinations of the physical parameters. First the standardized solution enthalpy table \mathbf{Z} can be calculated from the table containing the physical quantities (for example the solution enthalpy table in 6.1) with the equation

$$Z_{ij} = \frac{X_{ij} - \mu_j}{\sigma_j} \quad (3.76)$$

with μ_j and σ_j being the mean value and the standard deviation of column j . The PCC matrix can be determined directly through the equation

$$R_{ij} = \frac{1}{n-1} \mathbf{Z}^\dagger \mathbf{Z} \quad (3.77)$$

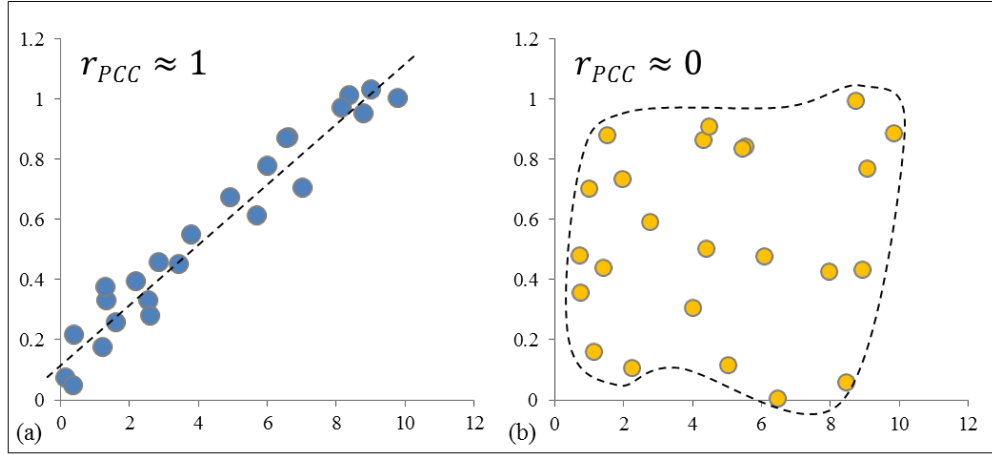


Figure 3.11: Schematic plots for the Pearson correlation coefficient for the two extreme cases (a) high correlation $r \approx 1$ and (b) no correlation $r \approx 0$. High anti-correlation corresponds to a value of $r \approx -1$.

in which n denotes the number of rows of matrix \mathbf{Z} .

3.7 Factor Analysis

The Factor Analysis is a method in the field of multivariate statistics in order to determine k latent/unobserved variables (so called factors) among m observed variables with the aim to reduce the dimension of the problem ($m > k$). All observed variables can then be determined (with a certain error bar) by a linear combination of these factors.

With the factors p_1, \dots, p_k and their loads a_{ij} each element of matrix \mathbf{Z} containing all observed variables in standardized form can be determined by

$$z_{ij} = p_{i1}a_{j1} + \dots + p_{jk}a_{jk} + \delta_i \quad (3.78)$$

with a certain error δ_i . The equation above is reformulated in matrix notation as

$$\mathbf{Z} = \mathbf{PA}^\dagger + \boldsymbol{\delta}. \quad (3.79)$$

Considering equation 3.79 and assuming the error to be negligible small the correlation matrix can be rewritten as

$$\mathbf{R} = \frac{1}{n-1}(\mathbf{PA}^\dagger)^\dagger(\mathbf{PA}^\dagger) \quad (3.80)$$

or

$$\mathbf{R} = \frac{1}{n-1}\mathbf{AP}^\dagger\mathbf{PA}^\dagger. \quad (3.81)$$

For non-correlated factors the multiplication gives the identity

$$\frac{1}{n-1}\mathbf{P}^\dagger\mathbf{P} = \mathbf{I}. \quad (3.82)$$

and for equation 3.81 follows

$$\mathbf{R} = \mathbf{A}\mathbf{A}^\dagger. \quad (3.83)$$

According to equation 3.83 the loading matrix \mathbf{A} can be determined through a singular value decomposition (svd) of the matrix \mathbf{R}

$$\mathbf{R} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\dagger = \mathbf{U}\mathbf{\Sigma}\mathbf{U}^\dagger = \mathbf{U}\mathbf{\Sigma}^{\frac{1}{2}} \left(\mathbf{U}\mathbf{\Sigma}^{\frac{1}{2}} \right)^\dagger. \quad (3.84)$$

in which $\mathbf{\Sigma}$ is a diagonal matrix with the eigenvalues of \mathbf{R} , and \mathbf{U} contains column-wise the corresponding eigenvectors. The loading matrix is then determined by

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}^{\frac{1}{2}}\mathbf{T}, \quad (3.85)$$

in which \mathbf{T} corresponds to any orthogonal rotation matrix. Starting with the identity matrix for \mathbf{T} (e.g. $\mathbf{T} = \mathbf{I}$) a particular solution of the loading matrix \mathbf{A} can be determined. The reduction to k factors appears by canceling out eigenvalues and corresponding eigenvectors according to different criteria (for example the Kaiser criterion only considers eigenvectors with eigenvalues larger than one because the corresponding factors covers more variance than a single observed variable did before). The reduced correlation matrix $\mathbf{R}_{\text{red}^k}$ can be determined by

$$\mathbf{R}_{\text{red}^k} = \mathbf{A}_{\mathbf{I}^k} (\mathbf{A}_{\mathbf{I}^k})^\dagger \quad (3.86)$$

The error due to the reduction can be determined by taking the difference of the original correlation matrix and the reduced correlation matrix

$$\boldsymbol{\delta} = \mathbf{R} - \mathbf{R}_{\text{red}^k}. \quad (3.87)$$

Equation 3.85 denotes that any orthogonal rotation \mathbf{T} of the particular solution is again a solution. A rotation allows a convenient visualization and interpretation of the factors. The rotation matrix \mathbf{T} is determined in such a way that the columns (e.g. sum of variances) of the squared loadings of matrix \mathbf{A} are maximized (the so called varimax method [118] is used here). An example is presented in the appendix.

Chapter 4

The workbench

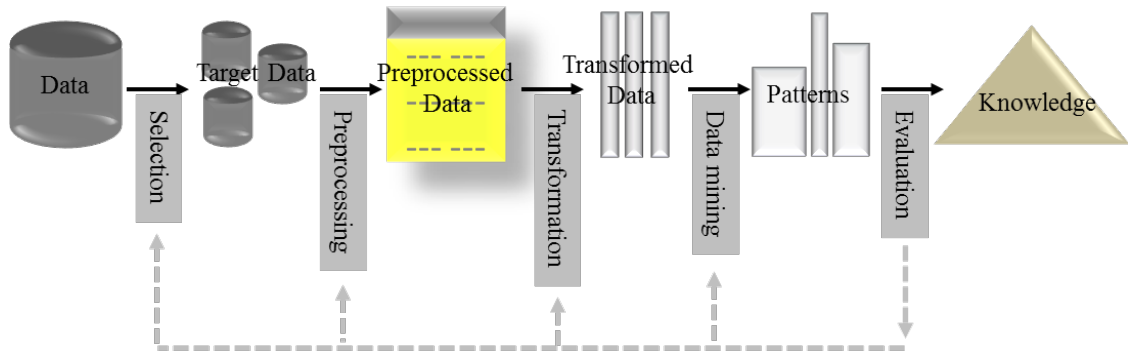


Figure 4.1: Figure adapted from [119]. Knowledge Discovery in Databases (KDD) process according to Fayyad *et al.*. The core of this process is the data-mining step, which needs an application of a specific algorithm in order to determine patterns from the data. The technical realization of this scheme is an in-house developed workbench.

In 1996 Fayyad *et al.* published their general concept of Knowledge Discovery in Databases [119, 120] as illustrated in figure 4.1. The whole KDD process is in the best case a fully computational process. Therefore, a workbench with several tools is required that can (i) handle each step of the full process in its complexity and can (ii) provide a user friendly communication between the computer generated data and its user. For this purpose, the development of algorithms for the most important steps within a workbench called pyCMW was a substantial part of the present thesis. We designed it such that the specific requirements of the KDD process are taken into account. The workbench is separated into different tools that are explained in the following.

4.1 Concepts

Front-end and Back-end

The KDD process requires full flexibility of the user to transform her/his data and at the time the chance to quickly visualize dependencies in order to identify possible pattern(s). We realized

that this is best fulfilled if a strict separation into a front-end and a back-end of the workbench is performed. In this context the front-end code contains the libraries, which are responsible for the generation of the graphical user interface (GUI). The back-end delivers all necessary workbench libraries needed for data creation, data analysis and data visualization (see figure 4.2). It is completely free of front-end libraries and can be directly used from a Python shell. To be more specific, a requirement for data visualization provided in the back-end is the “structure output”, i.e. to provide the atomic coordinates and atomic species on the command line. However, if needed, a call of a ‘showGUI’ on the command line, will switch to the front-end and will open a GUI for the specific tool, e.g. `AtomStructure.showGUI()`.

The front-end is using libraries from QT, which is a cross-platform framework. In this thesis we are focusing on the back-end, since this contains the important concepts in terms of KDD. It is constructed in such a way, that front-end tools, which are not specific for the materials science approaches addressed here, can be easily adapted to it.

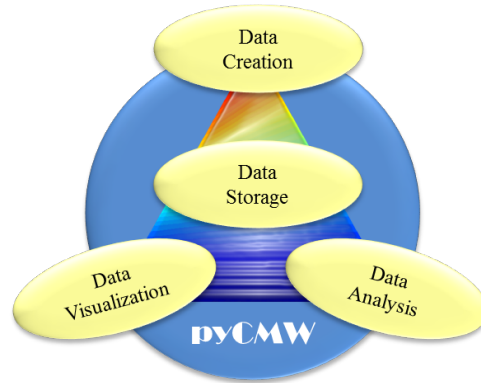


Figure 4.2: Tools of the developed workbench pyCMW (data creation, data visualization, data analysis, data storage). A strong separation between the front-end and back-end allows a general flexibility in the usage of the workbench.

Consideration of various external programs

The KDD process also requires a flexibility in the selection of tools for the generation of physical data from a given set of input parameters. A workbench in which the connection with only one DFT program is hard-coded would be insufficient. It is, therefore, a major feature of the pyCMW workbench that the communication with external programs is generic and supports various programs on different time and length scales simultaneously (see figure 3.1). In fact, if typical external programs used for atomistic simulations in materials science, such as VASP [121] and LAMMPS [122], being representatives for a DFT and a molecular dynamics code, are compared with one another, we find that from a physical point of view all information processed by them can be subdivided into two main modules: into the atomic **Structure** and into the **Hamiltonian**.

From a practical point of view, the communication with external programs happens via the program-specific input and output files. Despite the diversity of file structures, certain classes of

programs share identical physical data, so-called generic information. One example are the atomic coordinates that are intrinsically required for all kind of atomistic simulations. Other physical data like the total energies resulting from the calculation are not identical, but are of the same kind.

This general observation implied a certain structure during the implementation of the workbench classes. One of the cornerstones of pyCMW is that generic information is collected in one generic class (see figure 4.8) and used in the same way by all sub-classes. To make this approach computationally efficient, we use the concept of inheritance of object-oriented programming for this purpose.

4.2 Implementation

The mentioned subdivision into two main modules has been applied to our workbench as shown in figure 4.3. The highest class in the hierarchy is the **Project** class which needs at least one

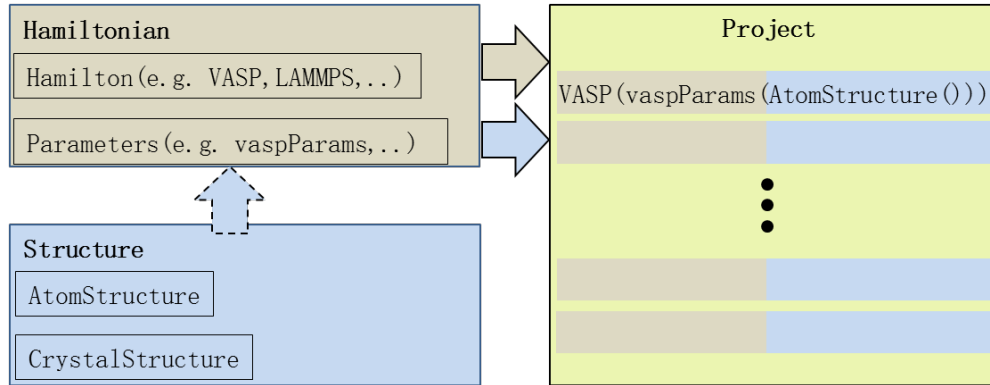


Figure 4.3: Subdivision of the modules and corresponding classes. A defined project can contain several jobs. A single job must contains an instance of the **Hamilton** class (e.g. **VASP**, **LAMMPS**,...), an instance of the corresponding **parameters** class (e.g. **vaspParams**, **lammpsParams**,...) and an instance of the **AtomStructure** class.

Hamilton instance as an input-parameter (e.g. **VASP**, **LAMMPS** or **KMC**). The **Hamilton** class needs a **parameters** instance and the **parameters** class an **AtomStructure** instance as input-parameters as shown in figure 4.3. It should be mentioned here that all instances can be generated during the program run or can be imported from existing so called HDF5 files (Hierarchical Data Format that will be discussed in section 4.3), which contain all necessary information for the specific instance. The mentioned classes are presented in more details in the following.

Structure classes: **CrystalStructure** and **AtomStructure**

The **Structure** module contains the **CrystalStructure** and the **AtomStructure** classes. The **CrystalStructure** defines the Bravais lattice, the Bravais basis, the lattice constants, and the chemical element of a perfect (and well defined) cell based on the fourteen Bravais structures and does not allow further modifications of the cell.

An example code is shown in figure 4.4. Line 1 generates an instance of a **CrystalStructure** class consisting of the element iron with a cubic Bravais lattice and a fcc basis which corresponds to setting four basis atoms. Line 2 generates a similar structure but with the element manganese in this case. For a "cubic" Bravais lattice which is a category for all cubic lattices the user has to choose one of the known three types: "fcc", "bcc" or "primitive". The **CrystalStructure** class

```
1 basis1 = CrystalStructure(element= 'Fe' , LatticeConstants=3.4,
                           BravaisLattice = 'cubic' , BravaisBasis= 'fcc' )
2 basis2 = CrystalStructure(element= 'Mn' , LatticeConstants=3.4,
                           BravaisLattice = 'cubic' , BravaisBasis= 'fcc' )
```

Figure 4.4: Generation of two **CrystalStructure** instances basis1 and basis2 corresponding to two separate fcc cells. At this level further modifications of the cells are forbidden.

delivers on purpose very limited functions. For further modifications the **CrystalStructure** must be transformed to an **AtomStructure** in which any modification on the atomic structure is allowed, which can be for example a shifting of the cell (line 6), the addition of magnetism (line 7), the shift of planes, replacing atoms etc. Alternatively, one can directly use the **AtomStructure**, in order to generate molecules for example. An example code which is a continuation of the previous code is shown in figure 4.5. Line 3 and 4 cast the previously generated **CrystalStructure** instances to **AtomStructure** instances in order to modify them. Line 8 produces a diamond cell (consisting of two fcc cells shifted with respect to each other). The sum in this line is an overloaded "+" operator resulting in a new **AtomStructure** instance. Further modifications on the diamond cell like increasing (27-fold) the cell, generating vacancies or adding an interstitial hydrogen atom (lines 9, 11 and 12) are performed. It should be mentioned, that the presented functions usually allow more input options as presented here. For example the "index=1" term in line 11 can be replaced by the term "element='Mn'" or in order to generate surface structures a command like "deleteAtoms(z > 2.)" will generate slabs. Line 14 stores the newly generated **AtomStructure** in a HDF5 file named **structure.h5file.h5**.

A generated **CrystalStructure** class must be transformed to an **AtomStructure** class since the **parameters** class of the **Hamiltonian** module (see figure 4.3) only allows an instance of the **AtomStructure** class as an input-parameter.

Hamiltonian classes: Hamilton and parameters

The **Hamiltonian** module contains two classes (see figure 4.3). The first class is the **parameters** class which is the hierarchically lower class in this module.

This class contains functions and variables that are influencing the specific **Hamiltonian**. Comparing the external programs VASP and LAMMPS common and non-common functionalities are observed. This allows a separation/extension of the **parameters** class and in a different naming of the inheriting **vaspParams** class and the **lammpsParams** class. However, generic information and functions (i.e. in common) of the programs are collected in the common parent class (**parameters**) and non-common information and functions in the corresponding child classes. This is shown in


```

3 basis1 = basis1.getAtomStructure()
4 basis2 = basis2.getAtomStructure()
5
6 basis1.shift(vec=[0.25, 0.25, 0.25])
7 basis1.addMagnetism(type='fm',magmom=3)
8 myAtomStruc = basis1 + basis2
9 myAtomStruc.repeat([3,3,3])
10
11 myAtomStruc.deleteAtom(index=1)
12 myAtomStruc.addAtom(element='H', coords=[0.5, 0.5, 0.5])
13
14 myAtomStruc.writePyTables(h5= 'structure_h5file.h5' )

```

Figure 4.5: Modification of the cells (e.g. shifting, adding magnetism, merging, cell increasing, deleting, adding (interstitial) atoms, etc.) is allowed only on the **AtomStructure** level. The **CrystalStructure** instances must be transformed to **AtomStructure** instances via the command `getAtomStructure()`.

figure 4.6.

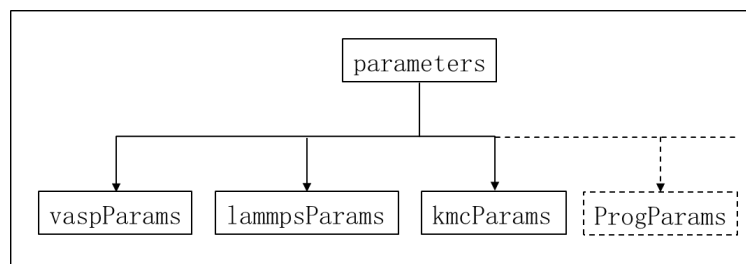


Figure 4.6: Class concept for the parameter classes. All parameter classes of external programs are inheriting from the parent class, namely **parameters**.

Handling and unifying the input and output of physical quantities of the external programs requires careful book-keeping due to the variability of input formats requested by the external programs. For example it can be the usage of different physical units in each external program (e.g., energy can be in Hartree, eV, Joule, etc.) or a clear definition of valid input values, which can be for example a data range or a specific data type. In a more complex case, input formats could be hierarchical, i.e. the usage of a given input parameter makes the usage of further input parameters necessary.

The program specific information about the possible flags, used units, input values, and the relation between them must be considered. This kind of information is stored in a HDF5 template-file called `vasp_default.h5` in the case of VASP (`lammps_default.h5` or `kmc_default.h5` respectively).

An example code for the usage of the **parameters** classes of the two different programs (VASP and LAMMPS) is shown in figure 4.7. In lines 15 and 20 instances of **vaspParams** and **lammpsParams**

```

15 param4Vasp = vaspParams(structure=myAtomStruc)
16 param4Vasp.setValue(name= 'SIGMA' , val=0.1)
17 param4Vasp.setValue(name= 'ENCUT' , val=30, unit= 'rydberg' )
18
19 strucFromFile = AtomStructure().loadPyTables(file_name= 'structure_h5file.h5' )
20 param4Lammps = lammpsParams(structure=strucFromFile)

```

Figure 4.7: Usage of the **parameters** classes: **vaspParams** and **lammpsParams**. An instance of an **AtomStructure** class is necessary as input.

are generated, respectively. Both instances need the **AtomStructure** as input. The **AtomStructure** instance `myAtomStruc` in line 15 was generated and modified in the lines mentioned before but it can be directly imported from an existent HDF5 file called `structure_h5file.h5` as shown in the lines 19 and 20. In the lines 16 and 17 modifications of VASP specific input flags are done, i.e. the width of the fermi smearing and the cut off energy for defining the plane wave basis set. Here the unit of the energy is given in rydberg but also other units (e.g. hartree, eV) can be used.

The two programs VASP and LAMMPS have different input and output formats but their corresponding **parameter** classes (i.e. **vaspParams** and **lammpsParams**) are using the same **AtomStructure** class as an input-parameter (lines 15 and 20). This indicates a major advantage of our workbench but will be more clear in the discussion of the next class, the **Hamilton** class.

The **Hamilton** class contains information and functions for the communication between our workbench (especially between the generated HDF5 files) and the external program specific input and output files. Common information/functions of the external programs VASP, LAMMPS and KMC are collected in a parent class called **genericHamilton** and non-common information/functions in the corresponding classes **VASP** class, **LAMMPS** class and **KMC** class. This is shown in figure 4.8. The **genericHamilton** class which is the highest parent class shares generic information and functions which are inherited and used in all three considered classes, whereas the **MDHamilton** class shares common information and functions just for VASP and LAMMPS Hamiltonians. Non-common information, such as, e.g., the naming of the input and output files as well as the information how to generate the specific input files or how to collect the output data are stored in these lowest classes. This structure allows a certain flexibility and makes especially the inclusion of not yet implemented atomic or electronic structure codes very easy. An example code for VASP and LAMMPS is presented below.

The **Hamilton** classes need an instance of the corresponding **parameters** classes. In the case of the **VASP** class it must be a **vaspParams** instance (`param4Vasp`) as shown in line 21 in figure 4.9, and for **LAMMPS** it is **lammpsParams** (`param4Lammps`) as shown in line 22. Predefined “scenarios” can be used in order to set predefined flags for the specific need. For the case shown in figure 4.9 a structure relaxation is aimed. Line 24 writes all the information about the structure and parameters into a HDF5 file. Line 25 generates the default LAMMPS input files and line 26 submits the job to a local machine/cluster. Assuming the job finished correctly line 28 collects the output of the LAMMPS job into the HDF5 file. The storage of the input and also the output information in the same file is challenging due to the large output but can be handled with HDF5 file formats. It is

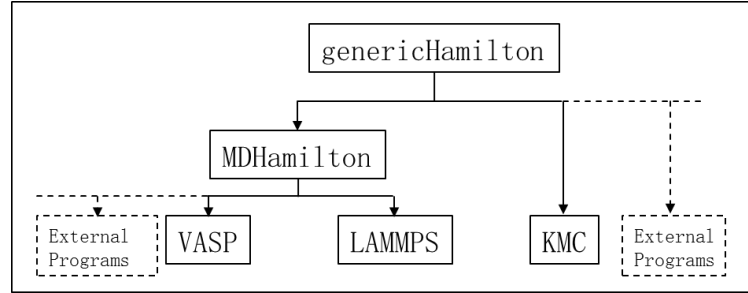


Figure 4.8: Class concept for the Hamiltonian classes. Generic information of different external programs are collected in the same class and inherited from the basis classes. This concept makes the inclusion of new programs into the workbench very easy.

```

21 vaspHam = VASP(param4Vasp)
22 lammpsHam = LAMMPS(param4Lammps, scenario= 'structure_relax' )
23
24 lammpsHam.writePyTables(h5= 'lammps_h5file.h5' )
25 lammpsHam.write()
26 lammpsHam.run(local)
27
28 lammpsHam.collect(h5= 'lammps_h5file.h5' )

```

Figure 4.9: Usage of the **Hamilton** classes: **vaspParams** and **lammpsParams**. An instance of an **AtomStructure** class is necessary as input.

obvious that the code from line 24 to line 28 also works for VASP having its own file formats.

```

29 strucFromLammpsRun = AtomStructure().loadPyTables(file_name= 'lammps_h5file.h5' )
30 param4Vasp = vaspParams(structure=strucFromLammpsRun)
31 ...

```

Figure 4.10: The key concept of pyCMW. A relaxed structure with LAMMPS can be directly used in a different program (VASP) for further calculations.

Starting a VASP calculation with the relaxed LAMMPS structure (line 29 and 30 in figure 4.10) is straight forward with this way of implementation. Since a computational “cheap” pre-relaxation of structures can be necessary in the present work (the incorporation of interstitial elements is breaking symmetries) which was the case for example in the work by Glensk for vacancy calculations with VASP [123]. In order to determine a ground state structure for cells with vacancies pre-relaxations with low convergence parameters were necessary by alternating calculations for the volume and for the ion positions. However, our concept described above allows alternatively to use a MD program (LAMMPS) in order to determine fast pre-relaxed structures for “expensive” electronic

VASP calculations.

Project class: project

The transformation of data in a KDD process typically involves higher-dimensional problems (e.g. pre-relaxation of the cell as described above, phonon calculations, fits to the Murnaghan equation of state, etc.) which require that several calculation steps are collected in the **project** class. In a simple case, this can be a loop over different lattice constants needed for a Murnaghan fit, but also more complex algorithm steps can be defined here. For the simple case of a Murnaghan fit the

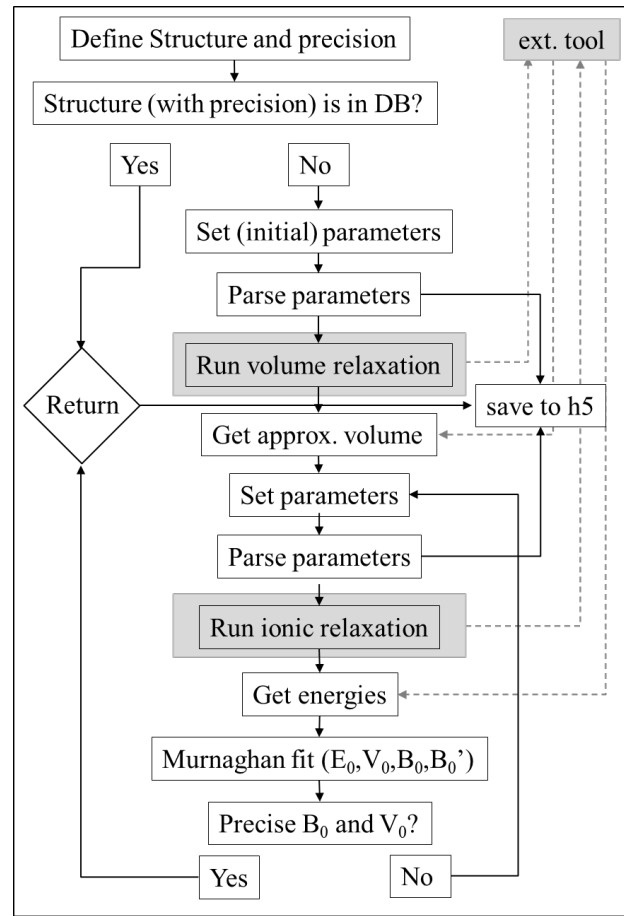


Figure 4.11: Algorithm for a fully automatized calculation of bulk modulus and equilibrium lattice constant of a defined structure. Gray shaded areas represent the steps performed in the external program. The steps of such an algorithm can be stored in a project class illustrated in figure 4.3.

algorithm is shown in figure 4.11. The whole process described by figure 4.11 can be performed fully automatically. This means no human interaction is needed and each of the steps will automatically start after the preceding step is finished. The corresponding example code is shown in figure 4.12. Line 1 generates the structure of interest. Line 3 defines the precision (i.e. the values for the bulk modulus and the equilibrium volume). The database is then asked for the structure with

the given precision (line 5) . If more than one item is found the results will be immediately plotted (line 8). If this is not the case convergence calculations will be started with the command

```

1 basis1 = CrystalStructure(element= 'Fe' , LatticeConstants=3.4,
2                             BravaisLattice= 'cubic' , BravaisBasis= 'fcc' ).getAtomStructure()
3 precision_bulk = 5 # GPa ; precision_vol = 1 # A^3
4 db = databaseAccess(dbname= 'mdb' , tablename= 'results_1' )
5 item = db.getItems(structure=basis1, hamiltonian= 'VASP' ,
6                     max_bulk=precision_bulk, max_vol=precision_vol, )
7 if item:
8     print item[ 'bulk modulus' ], item[ 'volume' ]
9 else:
10    myVASP = VASP(AtomStructure(basis1))
11    myProj = project(myVASP)
12    myProj.runConvergenceCalcs( [ 'B0' , 'V0' ],
13                                [precision_bulk, precision_vol], 'ENCUT' , 200, 'eV' )

```

Figure 4.12: The setup of an automatized bulk modulus and equilibrium volume convergence with respect to the cutoff energy. A schematic overview is given in figure 4.11.

`runConvergenceCalcs()` in line 12 (i.e. a function of the project class). In order to understand the necessity of the project class even in such a simple case we need a closer look to figure 4.13. It shows the code for the presented scheme (loop) in figure 4.11. An approximate volume is calculated in lines 2 and 3 (the member-function `getRelaxedVolume` is not explained here). Since the target is given as a list of two targets one has to consider both in the “if-case” in line 6 but for the case of simplicity we describe the code for a single target with the given initial target parameter (line 4). After the definition of a volume range and the number of data points (line 5) for the Murnaghan fit later in the code (line 10) the self consistent calculation (loop) is performed until the convergence condition(s) (current error is smaller than the targeted error) is (are) fulfilled (lines 6 to 16).

4.3 Tools of the workbench

As mentioned above (see figure 4.2) the workbench consists of four main tools required for KDD. All tools have in common that they can be directly used from the command line or with a graphical user interface.

Data storage

A fundamental feature of HT ab initio investigations is the production of huge amounts of data. We have therefore chosen HDF5 (Hierarchical Data Format) as an ideal candidate for the aimed data management (data storage, data compression, data sharing, data query, etc.) in pyCMW. Among several commercial and non-commercial software platforms (e.g. Java, MATLAB, ..), it is also supported by Python, since powerful interfaces and libraries (e.g., PyTables) have been developed by the Python community. HDF5 libraries are using chunked (bunches of certain length) datasets as atomic objects and the disk I/O is made in terms of complete chunks. This allows a very high

```

1 def runConvergenceCalcs(self, target, max_target_error, conv_param, init_val, unit):
2     self.getRelaxedVolume()
3     myVol = self.volumeRelaxed
4     target_diff = 10000.; target_last = 10000.; cur_conv_val = init_val
5     self.updateHam( 'Murnaghan', range=[myVol-0.1,myVol+0.1], pts=7)
6     if target_diff > max_target_error:
7         self.myHam.params.setParam(conv_param, value=cur_conv_val)
8         self.update()
9         self.submitProject()
10    myResultsDict= self.Murnaghan()
11    # returns a dictionary {B0:val, V0:val_v, E0: val_e, Bpr:val_b}
12    target_cur = myResultsDict[target]
13    target_diff = Abs(target_last - target_cur)
14    target_last = target_cur
15    cur_conv_val *= 1.1
16 else:
17     print target, target_last, target_diff

```

Figure 4.13: A function of the **project** class for automatized convergence calculations as shown in figure 4.11.

performance in terms of data indexing and data query [124]. The disadvantage is the high main memory usage for storing the index structure (so called B-tree) of the chunks. As mentioned above, PyTables is using HDF5 files and figure 4.14 shows the performance (time for complex queries) in comparison with a postgresQL database. It can be seen that complex queries on PyTables can be 10 times faster than for postgresQL tables and half disk space is used for the same amount of data. In addition one can see that for a large number of search hits the query time is increasing for postgresQL whereas pytables remain almost constant.

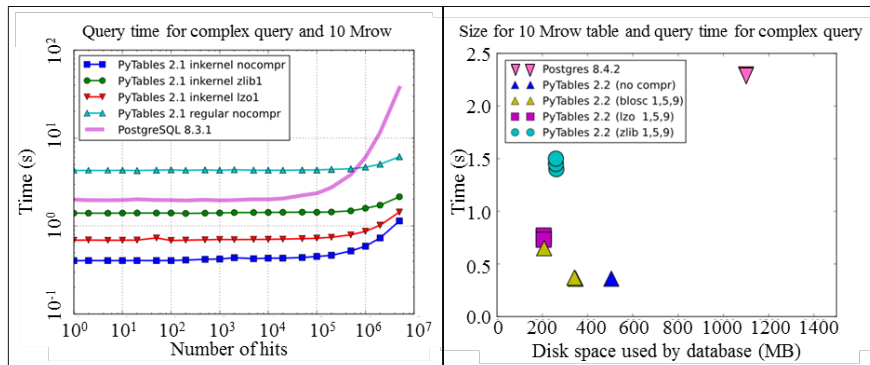


Figure 4.14: Performance test of the HDF5 format (by using PyTables) compared to postgresQL for a dataset of 10 million rows. The time for complex query for PyTables uncompressed data is 10 times faster than for postgresQL and only the half disk space is used. Copied from from [125].

The I/O of the different programs (VASP, LAMMPS, KMC, etc.) are stored in this unified file

format (HDF5). This allows a flexible information exchange between these programs (e.g. output of relaxed structure performed with LAMMPS can be directly used by VASP as input structure as shown above). It is very suitable to use the hierarchical form of the HDF5 file format in order to store the output of these programs which can also be ordered in a hierarchy as presented in figure 4.3. The figure 4.15 shows a different hierarchically filled HDF5 file. A project called “Murnaghan”

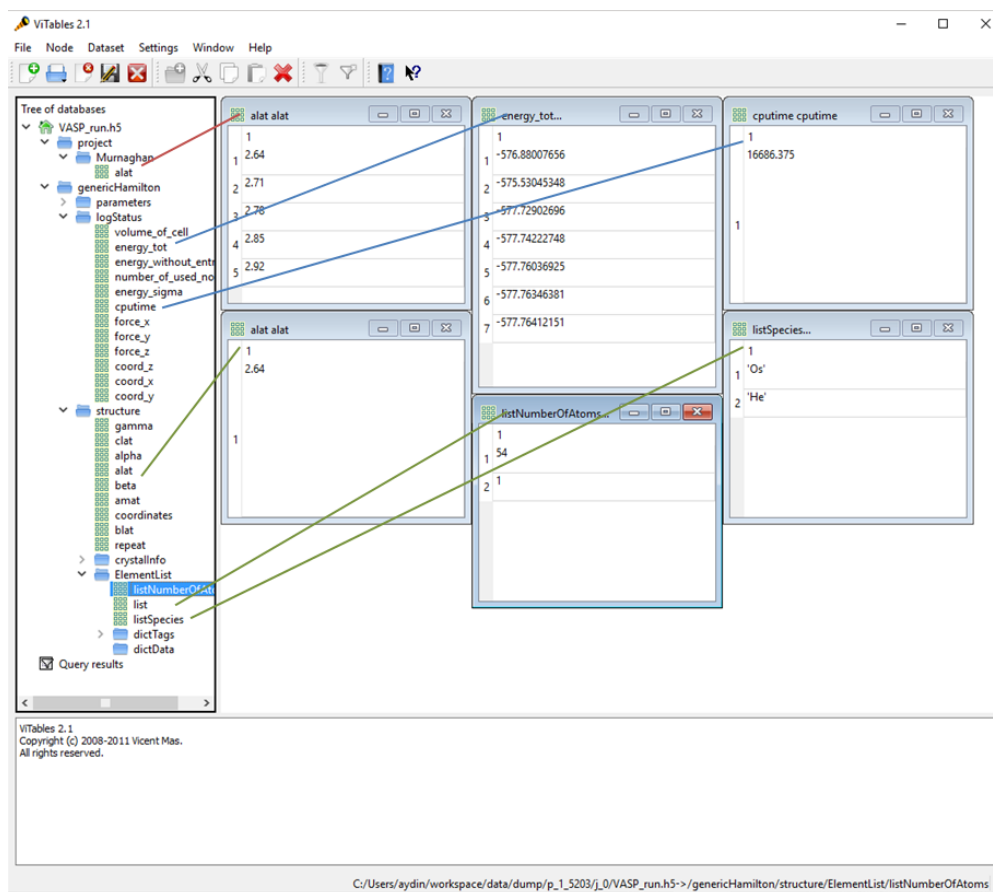


Figure 4.15: A hierarchically filled HDF5 file after the collection of a single “job” of a VASP Murnaghan-project containing five different lattice constants (red line). Expanded information about the energies after each ionic step and the total calculation time (blue lines) for the setup of 54 osmium and 1 helium atom at a lattice constant of 2.64 Å (green lines) is shown.

containing five different lattice constant calculations was submitted. The genericHamilton class is visualized as a folder containing sub-folders of the calculation results (logStatus), the hamilton (parameters), and the structure of the cell.

Data creation

The data creation is defined as a whole process beginning with the creation of the input data/jobs, via the calculation step with any external program, up to the collection of the output data as shown in the example above.

An intermediate automatic step is the parsing (translation) of the created instances of the classes to the specific requested input format of the external program (e.g. `vaspParams` will be parsed to INCAR, KPOINTS, POTCAR, POSCAR for VASP or `lammpsParams` to LAMMPS specific input files) or KMC, respectively. In addition a HDF5 file (default: `VASP_run.h5` containing full information of the input data will be created. This is shown in figure 4.16. At this stage a so called

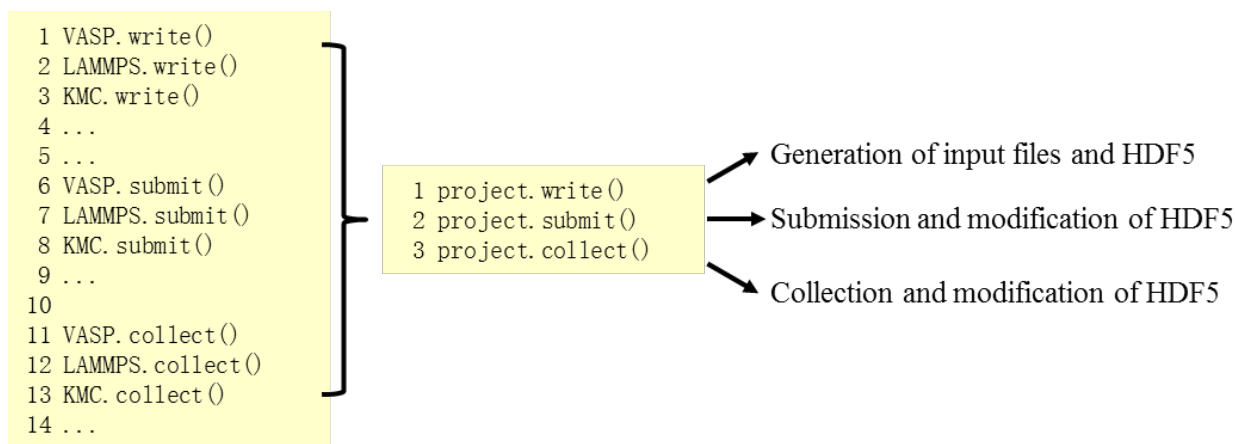


Figure 4.16: Schematic view of the same commands that can be applied for different external programs and will generate and read program specific input and output files.

“job” in a folder is created. In parallel all necessary information about the job (job status, job id, job path, ..) will be stored in a database. The database controls the running, finished, submitted jobs via a demon. Finished jobs are collected and the “hierarchical” filling of the HDF5 file with output data is performed automatically.

Data visualization

Data visualization in general means the presentation of input and output data to the user in two ways. The user can decide to visualize the data as output strings on the command line or use the front-end. For example the structure consisting of chemical elements, atomic coordinates, magnetization, etc. on the command line is presented as a multidimensional array. A `print structure.coordinates` call shows the structure coordinates, `print structure.elements` shows the elements of the structure, etc. The command `structure.showGUI()` opens a GUI, which presents the user different options from visualizing the atomic positions in an embedded Qt window to interacting on the structure directly (removing atoms, planes, changing elements, etc.). Data visualization includes also the visualization of finished calculations. Therefore, one or more HDF5 files, filled with output data, can be visualized on the command line or via a GUI (see figure 4.15).

Data analysis

Based on the database information and the HDF5 files output information (e.g. energies, volumes, forces) can be read in and analyzed via different data analysis tools. These tools are defined in the `genericHamilton` class if all external programs serve same physical information. For example a

Murnaghan fit needs cell volumes and corresponding energies as input which can be calculated with VASP or LAMMPS but not with KMC. The function for the Murnaghan fit will be defined in the common parent class, which is in this special case the **MDHamilton** class. The cluster analysis for example which is valid for all three programs is defined in the **genericHamilton** class. More complex analysis functions like vibronic frequencies of phonons, corresponding free energy calculations, or tools for the factor analysis introduced in section 3.7 are defined as separate classes or modules.

Job control

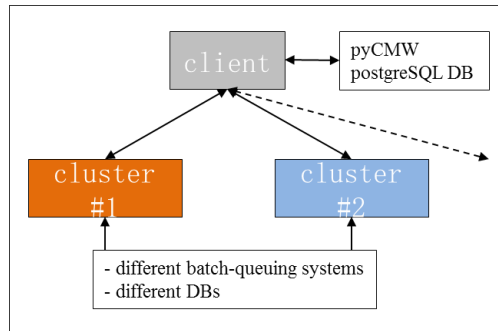


Figure 4.17: Schematic plot of the working principle of pyCMW. The clients database is separated from the different DB architectures and batch-queuing systems of the clusters.

In order to overcome the problem of different job management systems (i.e. batch-queuing systems) for different computer clusters, we establish our own postgresQL DB [126] which gets the filtered information coming from the different environments (e.g. cluster no.1, cluster no.2, etc. as shown in figure 4.17). The DB contains four tables: Two tables for current jobs (pending or running), a table of finished jobs and a table for error state jobs all involving information about project path, job path, project status, job status, etc.

Facing the problem that calculations can be performed on different locations (e.g. on local machines or on different clusters) having their own job management systems (e.g. SGE [127]) it is necessary to make the architecture of the workbench and its working environment independent of the several locations.

Chapter 5

Results I: Study of chemical trends

A key hypothesis of this thesis is that the solubility of interstitial elements is largely determined by physical properties of the host materials. The host materials under focus in this work are pure 3d, 4d, and 5d transition metals and their neighbors in the periodic table of elements (PTE) as shown in figure 5.1. The first part of this chapter contains, therefore, a study of these pure host systems only. We consider their ground state properties, provided they belong to the hcp, bcc, or fcc structures and have only a collinear magnetic order (e.g. nonmagnetic, ferro- or antiferromagnetic). In addition, we also consider all host elements under comparable conditions, for which we have chosen a non-magnetic fcc structure. Only in the second part, the occupation of the octahedral (o-site) or tetrahedral sites (t-site) of these host materials by light weight elements such as H, He, B, C, N, O, F, and Ne will be discussed. This study is focused on dilute concentrations of these elements, since this is a typical situation in experiments.

5.1 Study of host materials

The 4th, 5th, and 6th row elements of the PTE have been studied in hexagonal-close-packed (hcp), face-centered-cubic (fcc) or body-centered-cubic (bcc) crystal structures, assuming either a non-magnetic (nm) behavior or ferromagnetic (fm) configurations of their local moments. A bcc (collinear) anti-ferromagnetic structure was only used in the case of the host element chromium.

5.1.1 Self-consistent parameter optimization

Convergence of lattice constant and bulk modulus

For all ab initio calculations, it is an important prerequisite to ensure the convergence of the targeted physical data with respect to the numerical parameters employed in the program. Typically, this requires individual tests and adjustments by the user that are hardly compatible with a high-throughput data analysis.

We have therefore used the workbench pyCMW to automatize this procedure. The user defines a maximum error for a targeted physical property, as for example, the lattice constant or the bulk modulus of a crystal structure, and selects the set of considered convergence parameters, for example the cutoff radius for the plane wave basis functions or the number of k-points for the sampling of

I a											VIII a	
H 1s ¹											He 1s ²	
Hydrogen											Helium	
	II a	III a	IV a	V a	VI a	VII a						
		B 2s ² 2p ¹	C 2s ² 2p ²	N 2s ² 2p ³	O 2s ² 2p ⁴	F 2s ² 2p ⁵	Ne 2s ² 2p ⁶					
		Boron	Carbon	Nitrogen	Oxygen	Fluorine	Neon					
		p-band filling →					noble gases					
		Host materials										
		III b	IV b	V b	VI b	VII b	VIII b		I b	II b	III a	
4 th row	Ca 4s ²	Sc 3d ¹ 4s ²	Ti 3d ² 4s ²	V 3d ³ 4s ²	Cr 3d ⁴ 4s ¹	Mn 3d ⁵ 4s ²	Fe 3d ⁶ 4s ²	Co 3d ⁷ 4s ²	Ni 3d ⁸ 4s ²	Cu 3d ¹⁰ 4s ¹	Zn 3d ¹⁰ 4s ²	Ga 3d ¹⁰ 4s ² 4p ¹
	Calcium	Scandium	Titanium	Vanadium	Chromium	Manganese	Iron	Cobalt	Nickel	Copper	Zinc	Gallium
5 th row	Sr 5s ²	Y 4d ¹ 5s ²	Zr 4d ² 5s ²	Nb 4d ⁴ 5s ¹	Mo 4d ⁵ 5s ¹	Tc 4d ⁵ 5s ¹	Ru 4d ⁷ 5s ¹	Rh 3d ⁸ 5s ¹	Pd 4d ¹⁰	Ag 4d ¹⁰ 5s ¹	Cd 4d ¹⁰ 5s ²	In 4d ¹⁰ 5s ² 5p ¹
	Strontium	Yttrium	Zirconium	Niobium	Molybdenum	Technetium	Ruthenium	Rhodium	Palladium	Silver	Cadmium	Indium
6 th row	Ba 6s ²	La 5d ¹ 6s ²	Hf 4f ¹⁴ 5d ² 6s ²	Ta 4f ¹⁴ 5d ³ 6s ²	W 4f ¹⁴ 5d ⁴ 6s ²	Re 4f ¹⁴ 5d ⁵ 6s ²	Os 4f ¹⁴ 5d ⁶ 6s ²	Ir 4f ¹⁴ 5d ⁷ 6s ²	Pt 4f ¹⁴ 5d ⁹ 6s ¹	Au 4f ¹⁴ 5d ¹⁰ 6s ¹	Hg 4f ¹⁴ 5d ¹⁰ 6s ²	Tl 4f ¹⁴ 5d ¹⁰ 6s ² 6p ¹
	Barium	Lanthanum	Hafnium	Tantalum	Tungsten	Rhenium	Osmium	Iridium	Platinum	Gold	Mercury	Thallium
non TMs		early TMs			d-band filling →				late TMs		non TMs	

Figure 5.1: Interstitial elements and host materials that are studied in this thesis. The latter include mainly the 3d, 4d, and 5d transition metals. Since the study of chemical trends is extended to some neighboring metals, we typically refer to the host materials as 4th, 5th, and 6th row elements.

the Brillouin zone. The program then performs an automated value optimization of the mentioned convergence parameters with respect to the predefined error range. The challenge is to optimize all convergence parameters simultaneously. The principle algorithm, however, is similar to the one shown in figure 4.11.

For the purpose of the present study and as a compromise between the numerical efficiency for convergence tests and actual calculations, the self-consistent convergence scheme is not fine-tuned to perfectly optimize all numerical parameters. For example, to converge bulk modulus and equilibrium lattice constant, the cutoff radius is typically varied between the values 200, 300, and 400 eV (see figure 5.2). More sensitive dependencies such as the specific position of the lattice constants used for the evaluation of the Birch-Murnaghan fit [128] are not considered in the optimization procedure, though the developed code would in principle allow to do so.

c/a relaxation of hcp materials

As another example for automatizations implemented in the workbench, the relaxation of the c/a ratio of hcp crystal structures, is discussed next (see figure 5.3). This requirement for a shape relaxation makes the structure optimization a two-dimensional problem, what distinguishes hcp from the cubic cells (bcc and fcc). We have applied the following algorithm for this purpose.

1. Start with the ideal c/a ratio for hcp of 1.633 and find the equilibrium volume V_0 .
2. Relax the c/a ratio and keep the volume V_0 fixed.
3. Keep the c/a ratio fixed and relax and find the equilibrium volume V_1 .

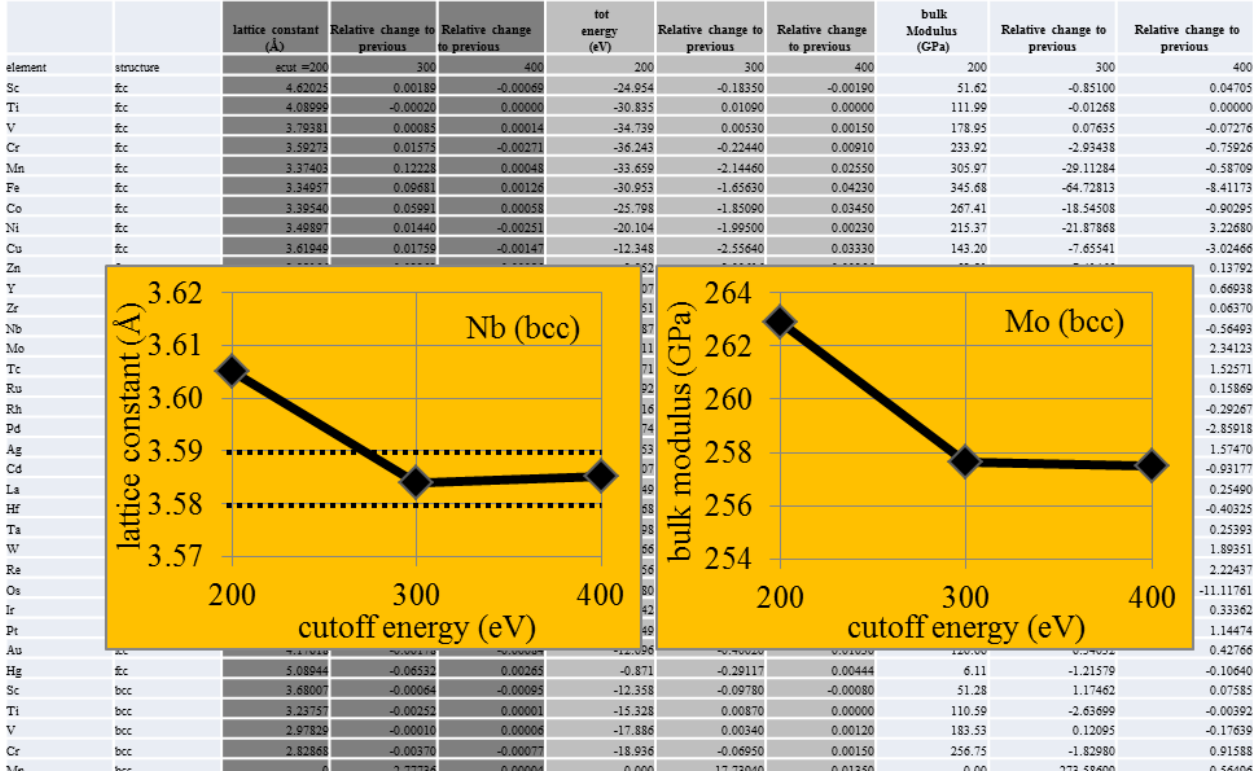


Figure 5.2: Efficient optimization of numerical parameters to ensure a pre-defined precision (dotted lines) of physical properties such as bulk modulus and the equilibrium lattice constant of a given element. In order to visualize only a fraction of the huge amount of generated data, a limited number of parameter choices are displayed.

4. Repeat the last two steps if the difference $V_1 - V_0$ is larger than a given confidence interval.

The c/a relaxation results for 4th, 5th, and 6th row elements are shown in Table 5.1 and demonstrate that overall a good agreement with experiments is achieved. A more meaningful benchmark for the c/a ratio algorithm is the comparison of the results with those obtained in a brute-force calculation of a whole mesh of c/a ratios and lattice constants. This comparison is graphically visualized in Figure 5.4 for the two elements Ti and Zn, which have qualitatively different deviations from the ideal c/a ratio. It clearly demonstrates the efficiency of the algorithm. A more quantitative comparison for further elements is shown in table 5.2. It demonstrates a qualitatively good agreement of the results achieved by the described algorithm with a calculation of a whole mesh of a and c lattice constants.

5.1.2 Chemical trends for host materials

In this section, we study the equilibrium volume, the bulk modulus, and the zero point energy for the elements in their respective ground state. On the one hand, the parameters of the host materials are required as reference values for solution enthalpies. On the other hand, the performance of HT investigations can already be tested and demonstrated: instead of discussing these quantities for

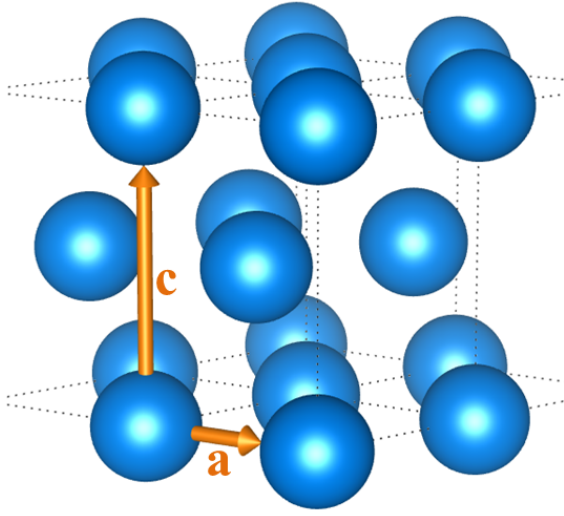


Figure 5.3: The lattice parameters (c and a) of an hcp supercell which are considered. According to a hard-spheres model the optimum c/a ratio is at approx. 1.633.

element	GGA-PBE	experiment
Sc	1.56	1.592
Ti	1.581	1.586
Co	1.617	1.622
Zn	1.877	1.861
Y	1.549	1.57
Zr	1.599	1.594
Tc	1.607	1.606
Ru	1.58	1.579
Cd	1.883	1.886
Hf	1.564	1.583
Re	1.617	1.616
Os	1.581	1.577
Tl	1.566	1.595

Table 5.1: Calculated c/a ratios in comparison with experiments [129]. Note that former are obtained at $T = 0$ K, whereas the latter are determined at room temperature.

a few elements only, the full set of all transition metals of the 4th, 5th and 6th row of the PTE is determined simultaneously. The large set of achieved results yield perfect statistics for testing correlation functions, data analysis strategies, and the relevance of the chosen crystal structure.

The cell structure and also magnetism might influence the trends of the bulk modulus and the equilibrium volume. Therefore we are interested in studying also the elements under comparable conditions. For this purpose we discuss the trends of the host elements additionally in non-magnetic (NM) fcc structure. The results are shown below.

The equilibrium volumes per atom and the bulk modulus for the host elements in their respective ground states are shown in figure 5.5. A clear chemical trend is observed for all studied elements. It can be seen that elements with half-filled d-band have smaller atomic volumes [130], i.e. smaller lattice constants, than the elements to the right and to the left of the periodic table. Further, it

element	mesh calc		algorithm		difference	
	alat	c/a-ratio	alat	c/a-ratio	Δ alat	Δ c/a-ratio
Ti	2.925	1.581	2.925	1.581	0	0
Co	2.501	1.624	2.495	1.617	-0.006	-0.007
Zn	2.666	1.87	2.66	1.877	-0.006	0.007
Cd	3.056	1.885	3.045	1.883	-0.011	-0.002
Tl	3.594	1.548	3.58	1.566	-0.014	0.018

Table 5.2: Results of the iterative algorithm for the lattice constant a (in Å) and the c/a ratio, compared to a whole mesh calculation.

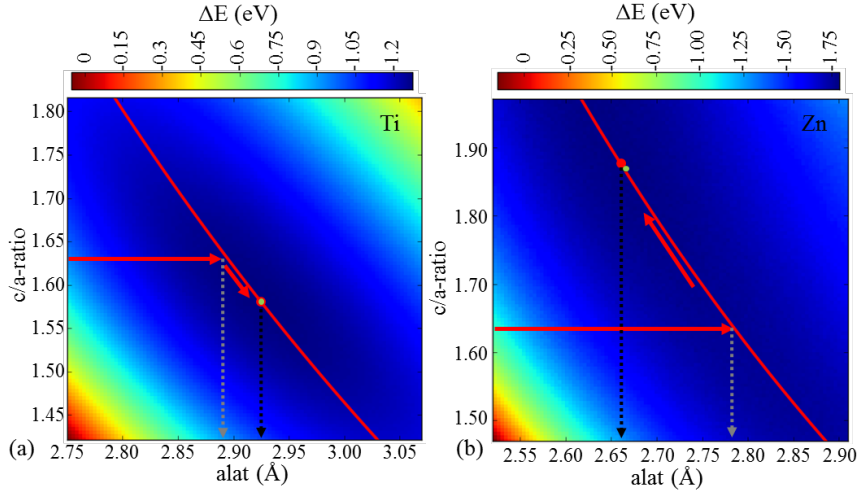


Figure 5.4: The relaxation algorithm (red arrows) compared to spline interpolated results achieved through a calculation of a large mesh (contour surface) for the host elements Ti (a) and Zn (b). The red line marks the constant volume curve $c/a = V_0/a^3$. The red dot marks the achieved equilibrium a value and c/a ratio after the first iteration with the suggested algorithm. The green dot is the minimum of the contour surface. The black dotted lines mark the lattice constant a and the energies after the application of the iteration scheme. The gray dotted lines mark the lattice constant a and the corresponding energies for a fixed c/a ratio without using the iteration scheme.

can for example be noticed that the 4th row elements have a smaller atomic volume, than those of the other two rows, whereas the differences are surprisingly small between the 5th and 6th row elements. This, so-called “lanthanide contraction” is known to be related to a reduced screening effect by 4f electrons and the resulting strong binding of the 6s electrons to the core [131, 132].

A reverse chemical trend is observed for the bulk modulus: Half-filled d-band elements have a higher bulk modulus than the elements to the right and to the left of the periodic table. These trends are well documented in the literature [130] and explained in terms of the underlying electronic structure. It is therefore not meaningful to repeat such an analysis here. Instead we provide a more detailed discussion on the trends for the ZPE, which have been less systematically investigated so far.

The zero point energy (ZPE) has been determined for all considered host materials using the method of the dynamical matrix, as explained in section 3.4. Trends for phonon energies have previously been performed to our knowledge only for subgroups of the transition metals [134, 135], as for example for the fcc metals [109]. This is mainly because the involved supercell calculations require, in comparison to the Murnaghan equation of state discussed above, orders of magnitude more computational resources. Furthermore, a careful convergence of phonon spectra with respect to numerical parameters is difficult to automatize in a HT investigation [109]. To avoid manual adjustments, we have for example removed those elements from the analysis, for which a straightforward application of the dynamical-matrix approach has lead to imaginary phonons.

The results are shown in figure 5.6a and reveal again clear trends: The phonon energies are

Bulk modulus ground state (GPa)											
Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga
17	54	115	184	179		170	220	191	136	68	
Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In
9		109	192	297	350	381	341	243	133		28
Ba	La	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Tl
11	38	93	164	255	286	293	250	165	87	37	
Bulk modulus experiment (GPa)											
Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga
17	57	110	160	160	120	170	180	180	140	70	
Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In
	41		170	230		220	380	180	100	42	
Ba	La	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Tl
9.6	28	110	200	310	370		320	230	220	25	43
Bulk modulus NM fcc (GPa)											
Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga
17	51	112	179	238	277	280	248	196	136	67	51
Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In
8	25	107	196	290	365	392	341	243	133	6	28
Ba	La	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Tl
11	39	93	162	245	299	301	250	165	87	41	36
Bulk modulus difference (NM fcc and ground state (%))											
Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga
—	-5.5	-3.6	-2.8	33.0	—	64.3	19.1	2.3	—	-1.4	—
Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In
—	-2.4	-3.3	-1.2	-3.9	0.1	-1.7	—	—	—	9.3	—
Ba	La	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Tl
-7.5	—	-5.3	2.0	-2.3	0.1	-1.8	—	—	—	—	4.1
V_0 difference (NM fcc and ground state (%))											
Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga
—	0.0	-0.1	3.4	0.6	—	-10.3	-5.1	-0.6	—	0.5	—
Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In
—	-1.5	-0.8	3.2	1.7	0.5	0.5	—	—	—	-1.3	—
Ba	La	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Tl
1.2	—	0.1	2.9	2.0	0.5	0.5	—	—	—	—	-1.6
ZPE difference (NM fcc and ground state (%))											
Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga
—	-12.7	-27.6	—	—	—	-7.1	-18.5	0.8	—	-29.3	—
Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In
—	-5.8	-20.8	—	—	-36.7	-35.4	—	—	—	-30.4	—
Ba	La	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Tl
-3.5	—	-26.6	-26.4	-43.4	-34.4	-36.1	—	—	—	—	-28.7

Table 5.3: Bulk modulus of 4th, 5th, and 6th row elements in their respective ground state (1st part) in comparison to experiments at room temperature (taken from [133]) (2nd part) and in NM fcc structure (3rd part). In the lower parts, a comparison of bulk modulus, equilibrium volume and ZPE for the elements in their respective ground state and in non-magnetic fcc structure is performed.

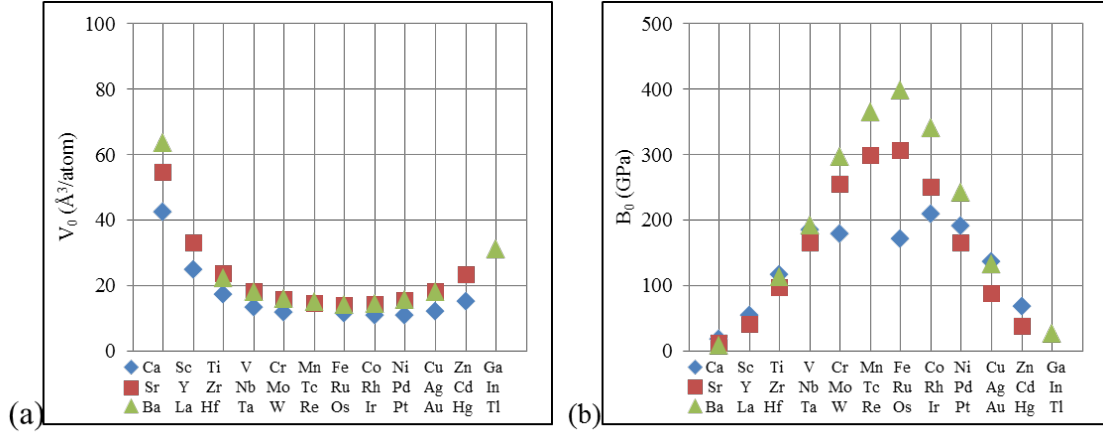


Figure 5.5: Equilibrium volume per atom (a) and bulk modulus (b) of 4th, 5th, and 6th row elements in their respective ground state. Chemical trends such as the maximum bulk modulus and minimum volume for half-filled d-bands can be observed. The ZPE energy of Nb is very high compared to the other host elements.

softer for lower rows in the periodic system and the ZPE is highest for elements with d-shells that are close to half filling. The trends are less systematic for the 3d metals close to half filling, where the magnetic order influences crystal structures and lattice constants and therefore also phonon energies. We have extended the phonon studies to quasiharmonic calculations, in order to determine for example thermal expansions systematically. However, instead of analyzing these thermodynamic properties, we would like to focus on the possible correlation of the observed trends, to prepare upcoming considerations for solution enthalpies.

Comparing the qualitative trends in the figures 5.5b and 5.6a yields the assumption that the ZPE is related to the bulk modulus of the material. This correlation is visualized in figure 5.6b and is physically evident. The host element niobium shows with approximately 112 meV for the ZPE a very high value compared to the neighboring elements in the PTE. Despite several tests, it remains unclear if a numerical or a physical reason is responsible for this deviation. The clarification is beyond the scope of the present work. For a more quantitative evaluation, we have used the statistical concepts introduced in section 3.6. Applied to all host materials, the two physical properties have a PCC of 0.56, which corresponds to a weak correlation only. This is partly related to the different trends among the individual d-electron shells. If one reduces the data set for the PCC determination to a single row in the periodic table, one obtains the significantly higher values 0.76, 0.75 and 0.93, for the 3d, 4d and 5d transition metals, respectively.

Constraint to NM fcc structures

The analysis of the ZPE and its correlation to the bulk modulus indicates that a straight-forward evaluation of all available data is often not meaningful. Magnetism, for example, destroyed chemical trends for the ZPE as discussed in the context of figure 5.6a. The reduction to certain subgroups improved the correlation coefficient. Furthermore, the crystal structures will also have an impact

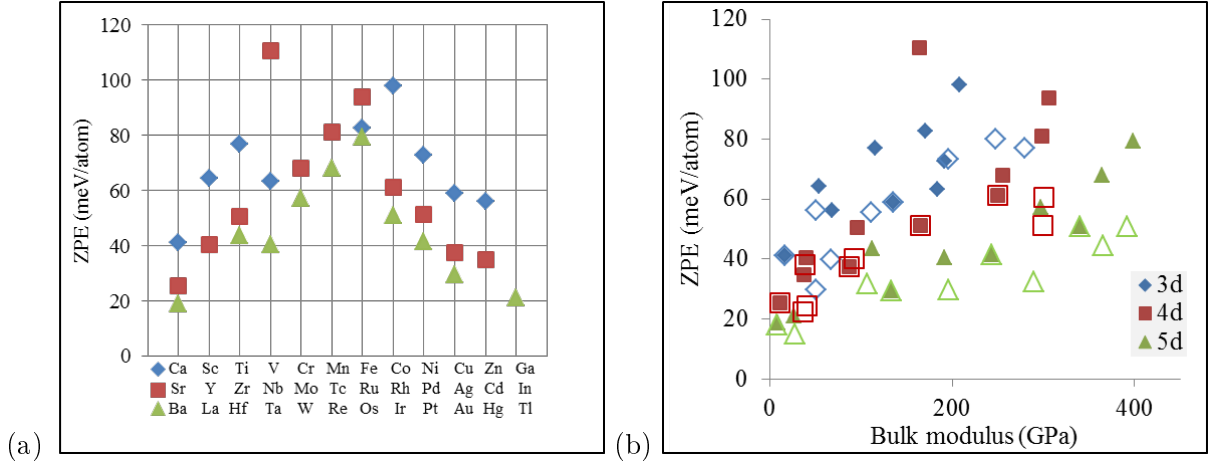


Figure 5.6: (a) Zero point energy for 4th, 5th, and 6th row elements (blue, red, green) in their respective ground state (filled markers). (b) Correlation of these zero point energies and the bulk modulus of the corresponding materials. Empty markers denote the correlation of the elements in NM fcc structure.

on chemical trends. It is obviously desirable to reduce the diversity of material properties to more comparable conditions, in order to focus the HT analysis on the chemistry of the considered materials. We have, therefore, repeated all calculations using the constraint of a non-magnetic fcc structure. The advantage of such an approach is meaningful. However, the comparison of both PCCs for the ZPE and the bulk modulus (GS vs NM fcc) reveal almost similar correlation coefficients for 3d, 4d, and 5d in GS 0.84, 0.95 and 0.93 and in NM fcc 0.90, 0.91 and 0.92 respectively. The increase of the PCC for the GS appeared because of canceling out the host elements which are not accessible in NM fcc structure. This is necessary in order to compare both structures since the PCC is very sensitive to small samples.

Table 5.3 provides the relevant numbers to evaluate the impact of constraining the magnetism and in particular the crystal structure on the bulk properties. In the upper part, absolute values for the elements in their respective ground state and in the NM fcc structure are compared for the case of the bulk modulus. The deviations are typically minor as compared to the discrepancy with the experimental values, caused by the approximation in the exchange-correlation functional. In the lower part of Table 5.3 relative changes of the bulk modulus, the equilibrium volume, and the ZPE due to the constraint are provided. The changes are largest for the magnetic materials and for the ZPE of non-fcc materials. This demonstrates that the employed constraints help to reduce scattering in the data, which is not primarily of chemical origin.

5.2 Comparison of hydrides, borides, carbides, and nitrides

In contrast to this, there are a number of experimental results for the high concentration phase for the interstitial elements H, B, C, and N available, i.e. hydrides (of the form MH or MH₂), borides, carbides and nitrides [136, 137]. For many transition metals such stable compounds can be formed,

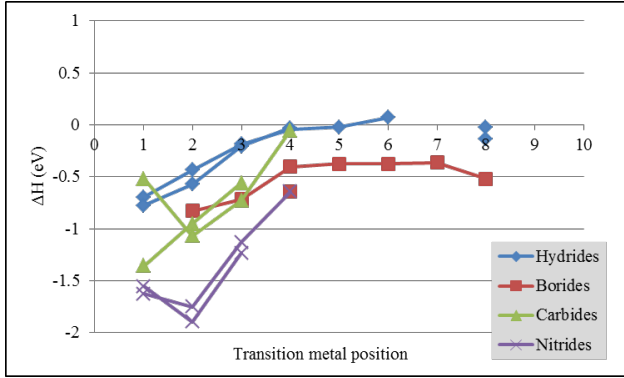


Figure 5.7: Experimental results for the enthalpy of formation for hydrides, borides, carbides, and nitrides of 3d and 4d TMs. The lines connect results for the same shell of TMs. The TM position corresponds to the data given in table 5.5.

(TM)H – (TM)N:	0.80
(TM)H – (TM)C:	0.63
(TM)H – (TM)B:	0.85
(TM)C – (TM)N:	0.62

Table 5.4: Correlation coefficients for the enthalpy of formations of the TM-interstitial compounds mentioned in table 5.5.

TM	(TM)B	(TM)C	(TM)N	(TM)H
Sc (1)	—	-131	-157	-67
Ti (2)	-80	-92	-169	-42
V (3)	-69	-54	-109	-18
Cr (4)	-39	—	-62	-4
Mn (5)	-36	—	—	-2.25
Fe (6)	-36	—	—	7
Co (7)	-35	—	—	0
Ni (8)	-50	—	—	-2
Y (1)	—	-50	-150	-75
Zr (2)	—	-103	-183	-55
Nb (3)	—	-70	-119	-20
Mo (4)	-62	-5	—	-3
Pd (8)	—	—	—	-13.3

Table 5.5: Enthalpy of formation (in kJ/mole atom) of experimental carbides, borides, nitrides (all taken from [136]), and hydrides (taken from [137]) of 3d and 4d transition metals. Their position in the PTE is given in brackets.

which is indicated by a negative enthalpy of formation. Table 5.5 provides the relevant numbers. It also shows that not all transition metals are forming stable compounds with interstitial elements.

When plotting these data in a diagram as a function of the d-band filling (transition metal position), see figure 5.7, one notices that the chemical trends for these high concentration phases are similar: Within a class of compounds (e.g. the nitrides) only the filling of the outer d-shell seems to be relevant for the formation enthalpy, not its number (3d, 4d, or 5d). The largest values are for all compounds observed for the TMs with d-shells close to half filling. The observations yield the interesting hypothesis that those transition metals that are very good hydride formers, also have a tendency to form borides, carbides or nitrides. Again the PCC is the best way to investigate such an assumption more quantitatively. We observe that the PCC of $r_{(\text{TM})\text{B},(\text{TM})\text{H}} = 0.85$ shows a high correlation between the hydrides and borides and less correlation between the hydrides and carbides $r_{(\text{TM})\text{C},(\text{TM})\text{H}} = 0.63$ (hydrides and nitrides respectively $r_{(\text{TM})\text{N},(\text{TM})\text{H}} = 0.80$). If an even better correlation coefficient than for the high concentration phases is achieved for the dilute limit, then a predictive formulation in the same spirit as Griessen’s predictive model for hydrogen (section 3.5) can be expected also for other interstitial elements. In particular, it will be interesting to realize, if the similarities between the impact of hydrogen and boron is also confirmed for the dilute limit.

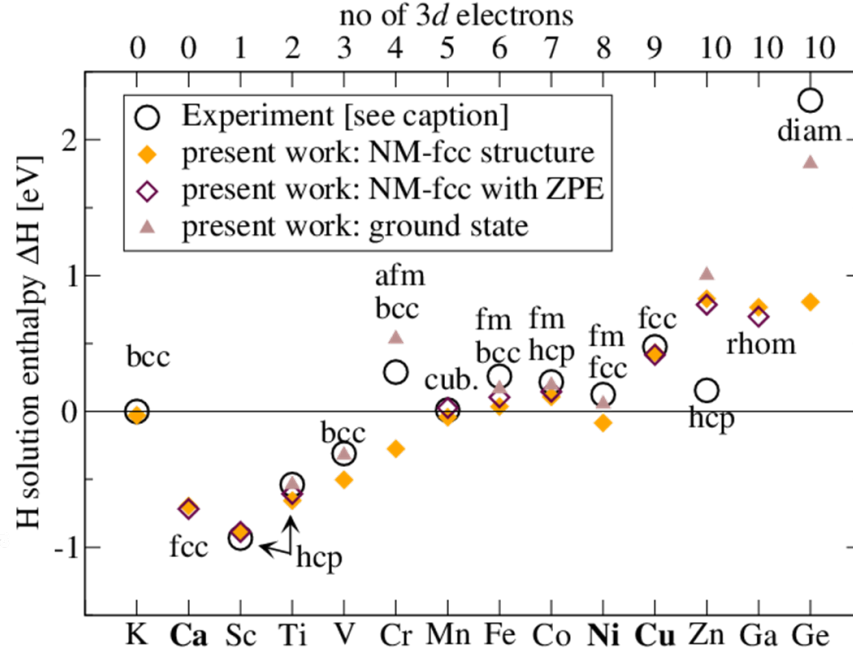


Figure 5.8: Solution enthalpies of H in 3d elements. Circles: Experimental data cited in Ref. [138]: K and Sc (Ref. [139]), Ti (Ref. [140]), V and Cu (Ref. [141]), Cr, Mn, Fe, and Co (Ref. [142]), Ni (Ref. [143]), Zn (Ref. [144]), and Ge (Ref. [145]). Filled diamonds: ab initio results for 3d elements assuming NM-fcc structures. Open diamonds: ab initio results including zero-point energy (ZPE) in NM fcc structure. Filled triangles: ab initio results for 3d elements in their ground state ($T = 0$ K) stable structure, as indicated by the label. The ab initio results always refer to the energetically most favorable interstitial site. Bold elements indicate where the fcc structure is the ground state. Figure taken from [146].

5.3 Study of hydrogen in 4th row elements

In order to exploit chemical trends in a high-throughput analysis, we first would like to get a physical understanding of the relevant mechanisms that determined interstitial solubilities. For this purpose, we limit our investigations to the interstitial element hydrogen in host materials of the 4th row in the PTE. In this section, the ab initio calculations are performed for non-magnetic (NM) fcc crystal structures which allow to separate chemical trends from properties of the structure and from magnetic interactions. The approach is also motivated by the interest in the local chemical environment in an fcc alloy (e.g. an austenitic steel) on the H solution enthalpy. The underlying assumption is that a local enrichment by impurity atoms does not change the lattice structure of the host matrix. The main findings of the study, which we have already been published [146] are summarized in the following.

5.3.1 Comparison of solution enthalpies with experiments

A comparison of the solution enthalpies determined by our ab initio approach and by experiment is shown in figure 5.8. One observes a nice agreement of theory and experiment data for the 3d TMs.

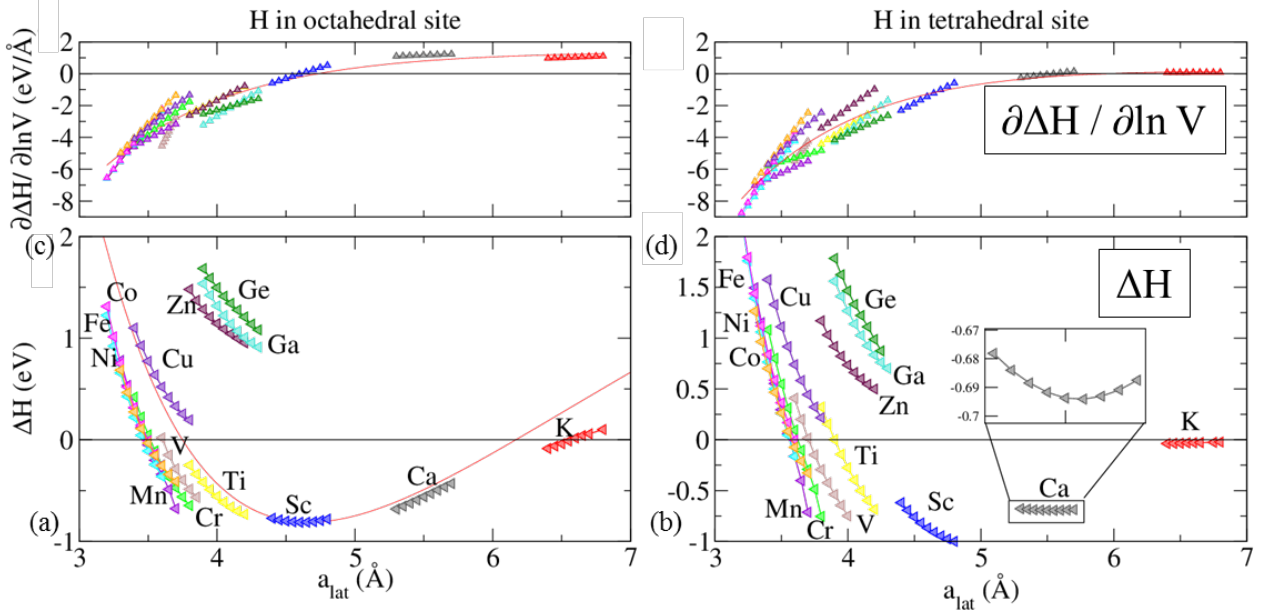


Figure 5.9: Solution enthalpy ΔH for: hydrogen in o-site (a) and in t-site (b). Furthermore the derivative of the solution enthalpy ΔH with respect to the logarithmic volume for hydrogen in the o-site (c) and hydrogen in the t-site (d) is shown. The red line indicates a fit of the resulting master curve of the simplified form $\Delta H(a_{\text{lat}}) = \alpha \exp(-a_{\text{lat}}) - \beta(1/a_{\text{lat}})$. The linear dependence of the derivative for a single host element is due to the 2nd order fit of the ΔH values. Copied from [146].

The figure further confirms that chemical trends become more apparent if all host materials are treated in the same structure and magnetic state rather than changing a complete set of parameters. If the ground state properties are determined, in particular Cr, Ga, and Ge show large deviations. Further, the figure demonstrates that corrections due to ZPE have only a small effect on the results.

Looking at chemical trends, we observe a clear dependence of the solution enthalpy on the filling of the outer d-shell. We can therefore assume that the 3d electrons have a systematic direct or indirect effect on the solution enthalpy. One candidate for such an indirect effect is the change of the lattice constant, which modifies the interstitial volume and therefore the solution enthalpy. In order to discriminate between direct and indirect effects the studies on lattice constants of the elements performed in Sec. 5.1.2, see figure 5.5, are helpful. They clearly indicate the correlation between the position of host elements in the PTE and the lattice constant. On the other hand, previous studies[116, 147] have indicated that apart from the chemical composition of the host materials the distance of the interstitial atom to the atoms of the matrix material is decisive for the H solubility. For example, the H solubility increases when alloying fcc Fe with Mn, since it expands the Fe lattice [147]. More generally, a proportionality between the interstitial Voronoi volume and the H solubility at octahedral or tetrahedral sites has been observed, while the chemical environment was of secondary importance.

5.3.2 Solution enthalpy as a function of the lattice constant

For the above mentioned reasons we systematically check the relevance of volume effects by plotting the solution enthalpies of figure 5.8 as function of the lattice constant (figure 5.9). Furthermore, we varied the volume for each of the host materials around the equilibrium lattice constant, corresponding to an hydrostatic compression and expansion of the lattice.

The solution enthalpy in Sievert's equation 3.39 is a function of pressure. However, a clear trend can be observed if the solution enthalpy is replotted as a function of the lattice constant. This is shown in figure 5.9 for the hydrogen solution enthalpy of the octahedral site (a) and its derivative (c) and separately the hydrogen solution enthalpy for the tetrahedral site (b) and its derivative (d). Two interesting effects appear:

- (A) A universal master curve appears for the solution enthalpy for both sites with a minimum solution enthalpy at a lattice constant of 4.6 Å for the octahedral and 5.5 Å for the tetrahedral site. This indicates a competition of attractive and repulsive interactions. An arbitrary chosen function of the form $\Delta H(a_{\text{lat}}) = \alpha \exp(-a_{\text{lat}}) - \beta(1/a_{\text{lat}})$ can be used in order to describe the volume dependence of the enthalpy curve shown as a red line in figure 5.12b
- (B) For the elements Cu and Fe (or Zn and Ti) which have almost a same lattice constants a constant shift in the enthalpy is observed.

In addition to the lattice constant our ab initio study in Sec. 5.1.2 also captured chemical trends for the bulk modulus for all TMs. Furthermore, the excess volume (equation 3.59) for H incorporation has been determined together with the solution enthalpy as a function of the lattice constant.

This allows to test the performance of the introduced analytic relation 3.60. First an axis transformation from the lattice constant (shown in figure 5.9a and b) to the logarithm of the volume is performed. Second the data is fitted to a third order polynomial in order to get the corresponding solution enthalpy derivative in order to compare it with the product of the pressure dependent bulk modulus $B(p)$ and the excess volume $V_{\text{H}}(p)$, which was determined through the Murnaghan equation of state. The results are shown in figure 5.10

The points are close to the analytic expression (solid red line), which allows the conclusion that the universal trend described in (A) is also applicable to the product of the pressure dependent bulk modulus and the excess volume. The remaining part of this section is dedicated to understanding the mentioned effects in (A) and (B).

5.3.3 Hard spheres model

If the host atoms are assumed to be hard spheres (with fixed radii) then the interstitial sites (o-site and t-site) must be sufficiently large enough to incorporate a hydrogen atom coming with its own fixed radius. The interstitial sites formed by hard spheres is presented in figure 5.11.

We assume that the fixed radii of the host atoms can be determined with the equilibrium lattice constant. It is then given as

$$r_{\text{host}}(a_{\text{lat}}) = a_{\text{lat}}/\sqrt{8}. \quad (5.1)$$

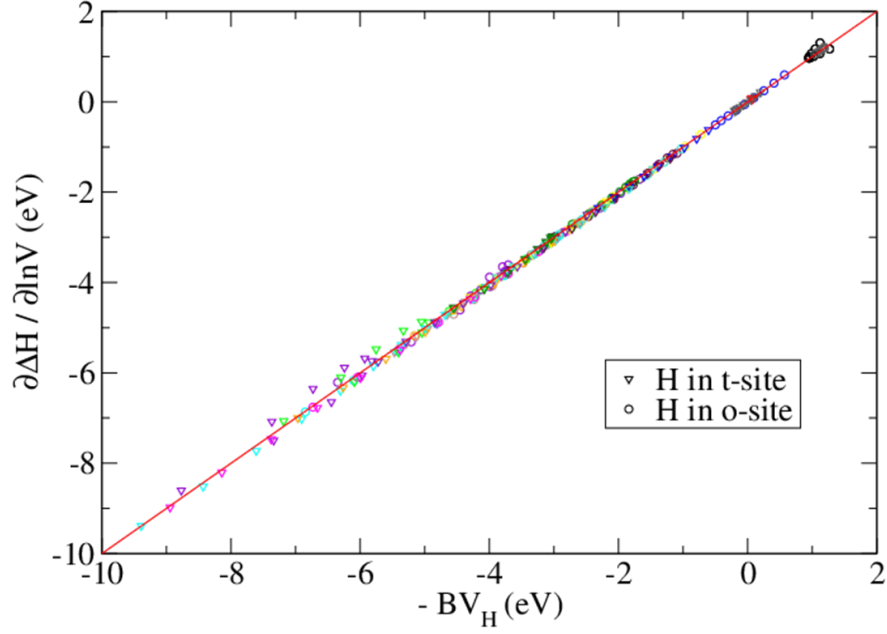


Figure 5.10: The derivative of the solution enthalpy as a function of the bulk modulus B multiplied by the excess volume V_H . The solid red line gives the analytic relation (equation 3.60). Copied from [146].

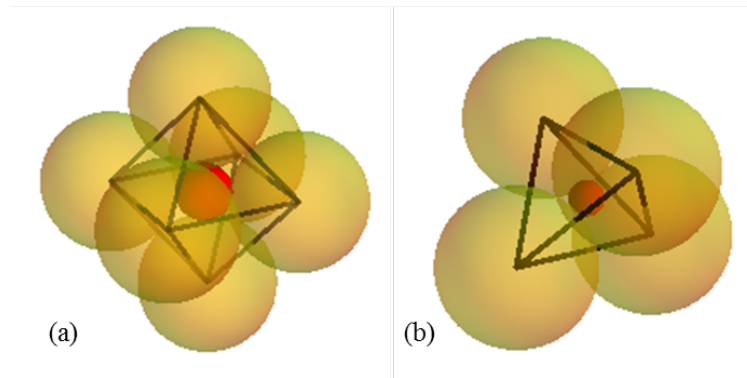


Figure 5.11: Hard spheres model: hydrogen (red) located in an o-site (a) at $a_{\text{lat}} = 4.6 \text{ \AA}$ and in a t-site (b) at $a_{\text{lat}} = 5.5 \text{ \AA}$ with host atoms (yellow) sitting at the octahedral (tetrahedral) corners. Taken from [146].

The radius of a sphere (e.g. hydrogen interstitial) which fits in an o-site is then given by

$$r_{\text{H}_{\text{octa}}}(a_{\text{lat}}) = a_{\text{lat}} \cdot (\sqrt{2} - 1)/\sqrt{8}, \quad (5.2)$$

and the interstitial radius in a t-site is given by

$$r_{\text{H}_{\text{tetra}}}(a_{\text{lat}}) = a_{\text{lat}} \cdot (\sqrt{6} - 2)/\sqrt{32}. \quad (5.3)$$

The host atom radii can be obtained $r_{\text{host}}(a_{\text{lat}}^{\text{octa}}) = 1.6 \text{ \AA}$ and $r_{\text{host}}(a_{\text{lat}}^{\text{tetra}}) = 1.9 \text{ \AA}$. We find for the H radii $r_{\text{H}_{\text{octa}}}(a_{\text{lat}}^{\text{octa}}) = 0.7 \text{ \AA}$ and $r_{\text{H}_{\text{tetra}}}(a_{\text{lat}}^{\text{tetra}}) = 0.4 \text{ \AA}$, i.e. both are significantly different. Interestingly, however, the host atom-H bond length in the o-site $r_{\text{host}}(a_{\text{lat}}^{\text{octa}}) + r_{\text{H}_{\text{octa}}}(a_{\text{lat}}^{\text{octa}})$ and in the t-site $r_{\text{host}}(a_{\text{lat}}^{\text{tetra}}) + r_{\text{H}_{\text{tetra}}}(a_{\text{lat}}^{\text{tetra}})$ is for both sites approximately 2.3 \AA , i.e. independent on the site.

A higher solution enthalpy is achieved for smaller lattice constants (e.g. more dense ion packing leads to smaller interstitial sites) which is in agreement with the studies of Syono *et al.* and Baranowski *et al.* reporting a hydrogen incompressibility in metals [148, 149]. However, the increase of the solution enthalpy when exceeding a critical lattice constant cannot be explained with such a simple model. Furthermore, the model predicts different H radii for both sites.

The universal master curve indicates that the observed o-site minimum at 4.6 \AA for the element Sc and the t-site minimum at 5.5 \AA for the host element Ca also appears for other host elements. We check this for the host elements Ca, Sc, Ti, Mn, Fe, Cu, Zn, Ga, and Ge. These host structures become unstable if their equilibrium lattice constant differs too much from the (o and t) site minima. Large tensile or compressive strain occurs. To avoid instability the host materials are forced to be ideal fcc structures (no relaxations allowed). The achieved o-site minimum for the elements Ca, Zn, and Ti correspond to that of Sc (a hydrogen radius of 0.7 \AA appears). The achieved optimum lattice constants for the elements Mn, Fe, and Cu is with approximately 4.1 \AA slightly smaller compared to 4.6 \AA . For the host elements Ga and Ge a larger optimum lattice constant is achieved at approximately 5.3 \AA . The observed range from 4.1 \AA to 5.3 \AA reveals a hard sphere radius for the hydrogen atom in the range of 0.6 to 0.77 \AA . In the t-site the resulting range of the interstitial hard sphere radius is between 0.35 to 0.44 \AA much smaller, which allows the conclusion that the optimum chemical interaction radius for hydrogen in the o-site (t-site) corresponds to a value between 0.6 and 0.77 \AA (0.35 and 0.44 \AA).

Ab initio studies on vacancy-hydrogen interactions in fcc Fe or fcc Ni revealed that hydrogen does not prefer the high symmetric vacancy center. The energetically more favorable interstitial site is slightly shifted in the $\langle 100 \rangle$ direction from the vacancy center [106, 150]. This leads to the question whether the H-M binding is strong enough to break the o-site symmetry for elements with lattice constants larger than 4.6 \AA . The hydrogen atom is displaced off-center at a distance of 2.3 \AA towards a single Ca atom (of two Ca atoms, respectively) with the result, that the H atom moves back into the o-site. The character of the interstitial configuration is not changing by applying compression or tension on the crystals which provides the basis for the universal master curve.

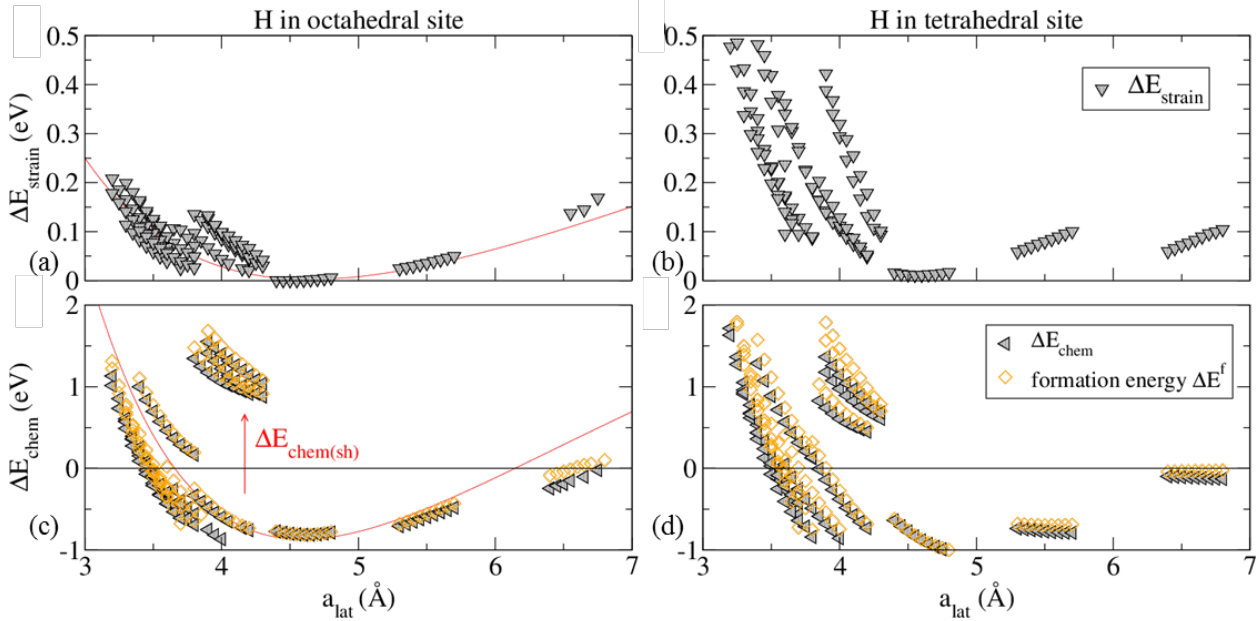


Figure 5.12: Separation of the solution enthalpy into a strain (a) and a chemical part (b) for o-sites. The strain energy is determined by relaxation effects of the host matrix whereas the chemical energy is determined by the interaction of the H atom with the matrix atoms. Both energy curves show an approximate master curve of the form $\tilde{\alpha} \exp(-a_{\text{lat}}) - \tilde{\beta}(1/a_{\text{lat}})$ (red). For H in the t-site the same separation into a strain (c) and a chemical part (d) is given. Copied from [146].

5.3.4 Separation of strain and chemical effects

The significant influence of the lattice constant on H solubility can have different physical origins. It is obvious to say that the lattice constant might be in correlation with atomic relaxation effects. The corresponding strain energy ΔE_{strain} can be decomposed from the solution energy by taking the difference energy between the two different configurations containing only the host matrix atoms (no hydrogen atoms):

$$\Delta E_{\text{strain}} = E_{\text{M}_N/\text{H}_1}^{\text{relaxed}} - E_{\text{M}_N}^{\text{ideal}}. \quad (5.4)$$

The first energy term in equation 5.4 corresponds to a crystal with relaxed host atoms positions due to hydrogen incorporation and removing the H atom and the second energy term to the ideal fcc crystal. The remaining part of the right-hand side of equation 3.44 is called chemical effect and determined by the equation

$$\Delta E_{\text{chem}} = E_{\text{M}_N\text{H}_1}^{\text{relaxed}} - E_{\text{M}_N/\text{H}_1}^{\text{relaxed}} - \frac{1}{2}E_{\text{H}_2}. \quad (5.5)$$

The separation into chemical effect and strain energy is shown in figure 5.12.

The relaxation energy, which is by definition always positive, shows a notably dependence on the lattice constant in the case of hydrogen in the o-site (figure 5.12a). As expected an increasing compressive strain energy is observed towards smaller lattice constants. Towards the opposite

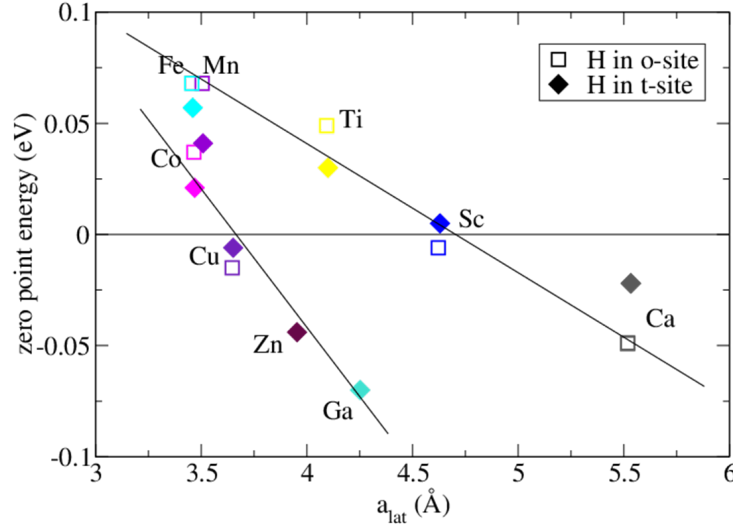


Figure 5.13: Effect of hydrogen zero point vibrations (ZPE) on the solution enthalpy. The ZPE are calculated within the harmonic approximation. Open squares: H in o-site, filled diamonds: H in t-site. Taken from [146].

limit of larger lattice constants an unexpected increase of strain (tension) energy is observed. In the compressive region the neighboring host atoms are pushed away and in the tensile region the neighboring host atoms are pulled towards the impurity. The strain energy vanishes at the optimum lattice constant of the master curve. The neighboring host atoms remain at their initial positions. There is no displacement of host atoms necessary to incorporate an hydrogen atom. For H in t-site (figure 5.12c) the compressive strain energy is approximately two times larger. The relaxation effects are stronger because of a smaller hydrogen-metal distance for identical lattice constants.

It is surprising that the strain energy vanishes around 4.6 Å for hydrogen situated in the t-site, whereas the optimum lattice constant for the tetrahedral solution enthalpy occurs around 5.5 Å. This observation implies that a simple hard spheres model which considers only 1st nearest neighbor atoms is limited in order to describe the solution of hydrogen. A model including 2nd, 3rd, etc. NN atoms is necessary. However, geometrical differences between the o-site and the t-site makes a direct comparison also difficult.

We employed a Voronoi construction and observed that the hydrogen atom in the o-site is completely surrounded by octahedral atoms whereas the hydrogen atom in the t-site has a contact surface to the 2nd NN atoms, which might lead to different chemical coordination effects for both sites. This is of importance since the chemical energy ΔE_{chem} is almost one order of magnitude larger than the strain energy ΔE_{strain} and mainly determines the solution enthalpy curve ΔH .

We therefore conclude that the shape of the universal master curve is mainly induced by the metal-hydrogen bonds and insignificantly by the hydrogen induced strain.

5.3.5 Zero point vibrations

As already outlined in section 3.4, zero point vibrations might be particularly relevant for a light weight element such as hydrogen. We have therefore studied the ZPE contribution for H solution enthalpy in different elements and different sites employing both methods introduced in section 3.4.

We have used the computationally fast Einstein method to calculate the volume dependent ZPE for hydrogen in Sc (o-site) and Ca (t-site). A compressive and respectively a tensile strain up to 5 % was arranged for both materials. The energy difference between the tensile and compressive region for both materials (and both sites) was less than 50 meV.

The results of the full dynamical matrix calculations is presented in figure 5.13. Imaginary frequencies were obtained for the elements V, Cr, Ni, and Ge, which indicates that these elements are unstable in an fcc structure. Nevertheless, the Einstein method still works for the mentioned elements. Imaginary frequencies for hydrogen in o-site was obtained for the host elements Zn and Ga.

The ZPE energy is increasing linearly towards half filled d-bands where it reaches a maximum. From half-filled d-bands towards completely filled bands the ZPE is decreasing linearly again. The chosen site (o and t) by hydrogen does not affect the ZPE trend. As mentioned previously for the strain energy in section 5.3.4 the ZPE almost vanishes at the identified lattice constant of 4.6 Å. The highest ZPE contribution to the solution enthalpy is less than 0.08 eV and will influence the solution enthalpy trend only moderately.

5.3.6 Chemical shift of ΔH

As shown in section 5.3.2 the slope of ΔH for hydrogen is independent on the specific element but for some host elements with similar lattice constants a chemical offset for ΔH was observed. This was shown in figure 5.9 (a and c). This leads to the conclusion that the chemical nature of the hydrogen-host interaction is directly responsible for the absolute value of the offset. In general we can say that elements with filled d-shells (e.g. Zn, Cu, Ga, and Ge) have a higher solution enthalpy than elements without filled 3d-shells (e.g. Ti, Fe, V).

In previous works by Griessen *et al.* [114, 116] the global electronic properties of the pure bulk system were identified to be the relevant physical parameters influencing the hydrogen solubility. Local chemical effects were mentioned but not considered in these works. However, it was known and discussed by Switendick *et al.* that local chemical effects, for example the electronic charge accumulation close to the shell of dissolved hydrogen, will appear [151]. We will focus in this section on these local effects and analyze the perturbation of hydrogen induced charge density differences (CDD) of the perturbed and unperturbed host materials. The CDD can be calculated with the equation

$$n(r)_{\text{CDD}} = n_{\text{MxH}}^{\text{relaxed}}(r) - n_{\text{Mx}}^{\text{relaxed}}(r) - n_{\text{H}}(r), \quad (5.6)$$

in which the first term corresponds to the electronic charge of the metal-hydrogen system and the second term to the relaxed pure metal system in which the host atom coordinates are identical to the host atom coordinates of the metal-hydrogen systems. The last term is the electronic charge of the hydrogen atom in vacuum. An example of a CDD isosurface is presented in figure 5.14

For all host elements an isotropic charge accumulation at the hydrogen position appeared. A more quantitative analysis is given by a 1D intersection as shown in figure 5.15.

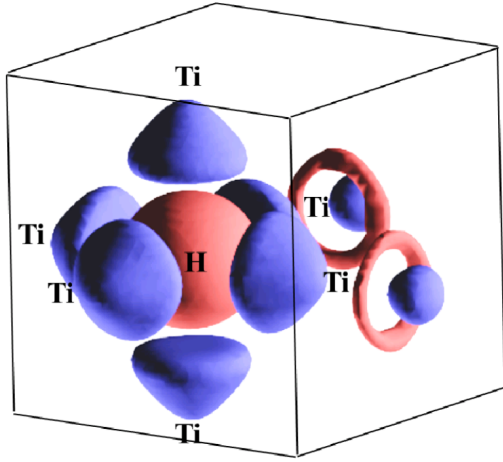


Figure 5.14: Isosurface for the charge density difference $n_{\text{CDD}}(r)$ using the example of Ti. The electron accumulating region (H position) is red coded and the donating region (host atoms) blue. For all elements the electronic accumulation at H is isotropic. Copied from [146].

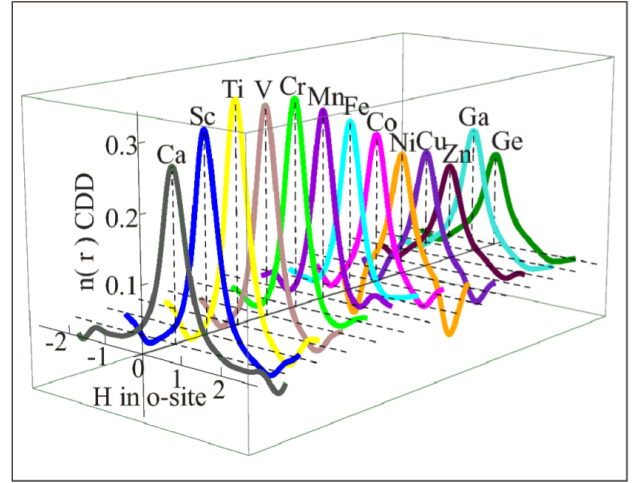


Figure 5.15: 1D intersections through the charge density difference between the systems with and without hydrogen for different elements. For H in an o-site a positive peak for all elements is seen and the electronic charge is increased. Copied from [146].

It is evident that the charge maximum observed for all elements in figure 5.15 is due to the attractive Coulomb interaction of the hydrogen proton. However, the amount of charge accumulation depends on the specific host element. Early elements of the PTE with a low number of 3d electrons like Ca, Sc, and Ti have higher perturbation of the electronic charge whereas late elements of the PTE like Cu, Zn, Ge with closed 3d shells show lower perturbation of the electronic induced by hydrogen. It can be understood within a simple bonding picture (*anion model*): hydrogen is binding an extra electron close to its 1s shell. An energy minimization will appear. Electron donating host elements with filled d-shells have a lower tendency to donate electrons than host elements with incompletely filled 3d shells.

In addition to the *anion model* the electron (coming from H) will be added at the Fermi-level which will lead to a repulsive electron-electron Coulomb interaction. Compared to early elements in the PTE the late elements of the PTE with filled 3d shells have a higher density at the Fermi-level. An increase of ΔH is the consequence (*proton model*) [151]. Nevertheless analyzing the DOS plot of the metal-hydrogen systems a remarkable change of the densities at the Fermi level were not observed.

However, based on the presented results above (figure 5.15) the solution energy increases with the CDD at the hydrogen position. In addition an occupation with two electrons of the s-like states was observed for almost all elements. For dilute hydrogen it stands to reason that the *anion model* seems to be a much more appropriate. A more quantitative analysis is revealed if ΔE_{chem} is plotted against the spherically integrated CDD (figure 5.15). This is shown in figure 5.16b.

The sign conversion of the CDD defines the radius of the integration-sphere. A clear correlation between the CDD and a reduced ΔE_{chem} is observed. With the ability of the host to donate electrons

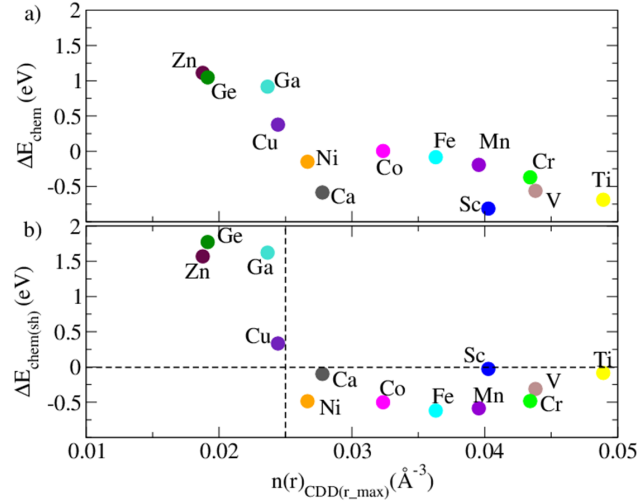


Figure 5.16: Chemical effect as a function of the spherically integrated CDD (see text): (a) The chemical effect vs the integrated CDD. Early elements of the transition series (Sc, Ti, V, Cr, Mn) show larger polarization effects. Such screening effects lower the chemical offset and simultaneously increase H solubility. (b) The element dependent chemical shift relative to the master curve vs the integrated CDD. The plot reveals a critical value of CDD for the elements, which separates the studied elements in two classes. Taken from [146].

the hydrogen solubility increases. If the energy difference between ΔE_{chem} and the master curve in figure 5.12 which we define as $\Delta E_{\text{chem}(sh)}$ is plotted against the spherically integrated CDD the correlation becomes more clear. A critical value of 0.025 electrons/ \AA^3 for the CDD separates the elements into two classes: All transition metals with not filled d-shells have a CDD larger than the critical value are below the master curve and on the other hand the host elements Cu, Zn, Ga, and Ge in which hydrogen is less soluble are below the critical CDD value.

5.3.7 Comparison with semi-empirical models

In this subsection we will use our ab initio data to assess the physical terms ΔE , W_d , and R_j of Griessen's semi-empirical model which was introduced in section 3.5.

According to Griessen *et al.* the solution enthalpy of hydrogen in a transition metal ΔH can be predicted with only the mentioned three physical properties of the pure bulk by using the equation 3.74. The host element affects with ΔE , W_d , and R_j three quantities simultaneously. In order to evaluate the validity of ΔH dependence on only one quantity, two of these quantities are taken for granted. The equation 3.74 can be rearranged as e.g. in the case of ΔE

$$\frac{\Delta H - \beta}{\alpha W_d^{1/2} \sum R_j^{-4}} = \Delta E \quad (5.7)$$

(respectively rearranging the equation for W_d and R_j).

The fitted parameters have the values $\alpha = 0.13 (\text{eV/atom})(\text{eV}^{-3/2} \text{\AA}^4)$ and $\beta = -1.07 \text{ eV/atom}$. As predicted by Griessen for the transition metals the values lie on the black line. Therefore, a

perfect correlation with the ΔE_{F-s} parameter is observed. The model can even be extended to Ga and Ge but is not working for the elements K, Ca, and Zn (see figure 5.17a).

The analysis of the parameter of the hydrogen-metal distance R_j reveals an almost same behavior (figure 5.17b). The model works for almost all transition metals and can be extended to the elements Ga and Ge. In section A the correlation between the d-bandwidth W_d and the rearranged equation 5.7 is shown.

We can conclude that the pure host parameters ΔE_{F-s} , W_d , and R_j which can easily be calculated with ab initio methods are convenient descriptors to predict hydrogen solution energies in 3d transition metals. The decisive question for the remainder of this work is, if the same concept can be extended to other host materials and/or interstitial elements or if even better models exist for these cases.

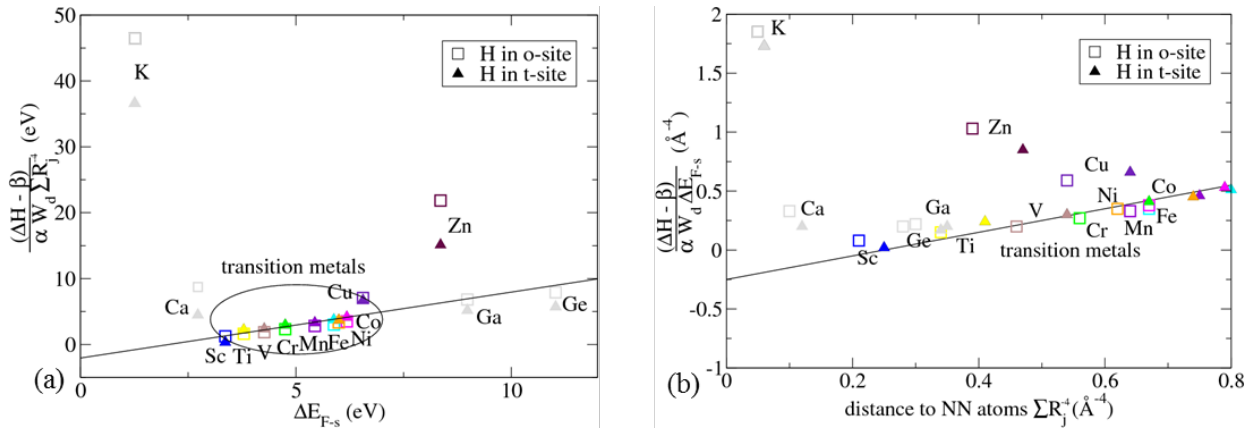


Figure 5.17: Correlation between Griessen's approach[114] for the H solution enthalpy the parameters (a) ΔE_{F-s} , (b) $\sum_j R_j$ and W_d (not shown here) as obtained from our ab initio data. Copied from [146].

5.4 Hydrogen solution enthalpy in 5th and 6th row elements

The investigation in the previous section was on purpose restricted to the single interstitial element hydrogen and host materials from the 4th row of the PTE. In this way, we achieved valuable insights into physical rules that determine the hydrogen solubility. In the following we shall extend the database to host materials of the 5th and 6th row of the PTE, repeating some of the previously performed investigations.

5.4.1 Hydrogen solution enthalpy in ground state structure

In this subsection the ab initio determined solution enthalpy of H in the ground state of the 5th and 6th row elements is compared to experimental data. In a similar manner as for the 4th row host elements, the results are summarized in figure 5.18. A comparison with figure 5.8 reveals that key observations are reproduced: First, it can be noticed that the DFT results show an overall nice

agreement with experimental data. Quantitative differences of about 0.5 eV are observed for the host materials Mo, Ba, and Au, but this does not influence the chemical trends.

Second, we observe the same trend for the dependence of the hydrogen solubility on the position of the TM in the PTE: Similar to the 3d trend in section 5.3 a negative ΔH is observed for hydrogen in the early elements in the 4d and 5d TM series. They tend to be hydride forming elements. By filling of the outer d-band the solution enthalpy is increasing and the incorporation of hydrogen becomes less likely. Even deviations from this trend, like the increased value for the VIb column (Cr, Mo, W) and the dip for the last VIIIb column (Ni, Pd, Pt) are reproduced for all three considered rows of the PET.

Third, zero-point vibrations have again been considered. They also have for the 5th and 6th row elements a noticeable, but small effect. In some cases, there consideration improves the comparison with experiment, but overall the chemical trends are not affected by ZPE corrections.

Therefore, we can expect that most of the results obtained in the detailed analysis of section 5.3 do equally apply to the transition metals considered in the present section.

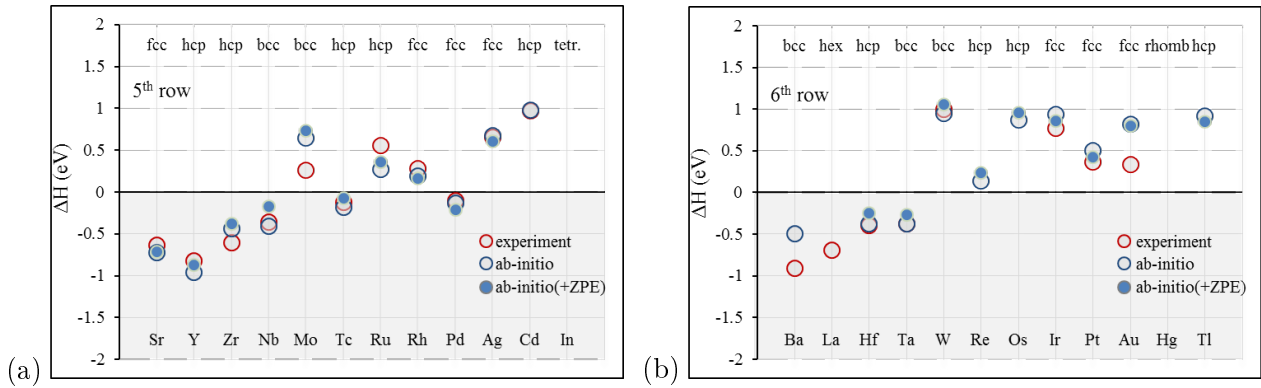


Figure 5.18: Solution enthalpy of hydrogen in (a) 5th and (b) 6th row elements of the PTE. Red circles corresponds to experimental data taken from references [138–145]. Blue open circles corresponds to ab initio data and the filled blue circles included zero point energy. An overall good agreement with experimental data can be seen.

Motivated by the above mentioned similarities, we also followed the philosophy of section 5.3 and assumed all 5th and 6th row host elements to be in a NM fcc structure, even if this structure is not the stable ground state. This constraint is applied in order to focus on chemical trends, while ensuring otherwise comparable conditions.

The effect of the structure on the hydrogen solubility is presented and compared in table 5.6. A general observation is that the constraint to the fcc structure reduces the solution enthalpy and therewith increases the hydrogen solubility. The reduction is particularly strong for some of the elements like Mo that have a bcc ground state structure – indicating again that a more open structure does not imply a better hydrogen solubility. Exception from this observation are small and limited to the elements Tc, Re, and Cd, which have all a hcp ground state structure, i.e. a structure that is very similar to the fcc structure considered here.

By increasing the row number (i.e. from 4th via 5th to 6th row elements) the solubility of H is slightly decreasing. The effect is stronger for host elements with more than half-filled d-shells,

ΔH of H in NM fcc vs GS (eV)												
element	Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga
NM-fcc	-0.533	-0.803	-0.651	-0.463	-0.274	-0.036	0.062	0.112	-0.078	0.429	1.186	0.997
GS	-0.533	-0.759	-0.447	-0.157	—	—	0.329	0.208	0.070	0.429	1.290	—
element	Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In
NM-fcc	-0.466	-0.730	-0.626	-0.456	-0.277	-0.059	0.134	0.190	-0.129	0.674	1.231	0.973
GS	-0.466	-0.702	-0.400	-0.128	0.936	-0.184	0.274	0.190	-0.129	0.674	1.197	—
element	Ba	La	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Tl
NM-fcc	-0.444	-0.774	-0.584	-0.376	-0.098	0.400	0.714	0.938	0.566	1.017	1.701	0.819
GS	—	—	-0.248	-0.052	1.359	0.133	0.869	0.938	0.566	1.017	—	0.916

Table 5.6: Octahedral solution enthalpy of hydrogen in non-magnetic fcc host elements compared to the solution enthalpy in their ground state structure.

i.e. elements with smaller equilibrium volume (compare figure 5.5). Furthermore, it is stronger for the change from 5th to 6th row, than from 4th to 5th row. These observations are related to the universal volume trend.

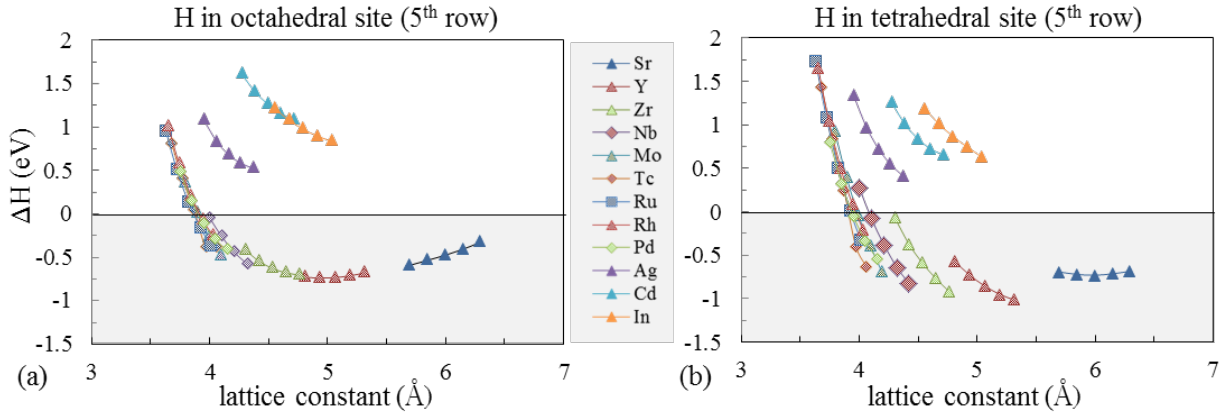


Figure 5.19: Solution enthalpy of hydrogen in 5th row elements in the octahedral (left) and in the tetrahedral (right) site.

This universal behavior was the most interesting result for the 4th row elements constraint to NM fcc structures, see figure 5.9. It is even more remarkable that exactly the same behavior is also achieved for the 5th and 6th row host materials. We show this result in figure 5.19 only for the 5th row elements, whereas a comparison for all investigated materials is provided in figure B.1 in the appendix. One can clearly see that the shape of the universal curve is for all rows of host materials almost the same. This becomes even clearer when comparing the derivatives in figure B.2. Even the absolute minimum along the universal curve is with approx. -0.8 eV always the same. Only the optimum bonding distance in the octahedral (tetrahedral) site is slightly increasing for the 5th row elements to approx. 5 Å (6 Å, respectively) and 6th row elements to approx. 5.5 Å in the octahedral site. H in the tetrahedral site of the 6th row elements does not show a minimum up to

a lattice constant of 7 Å. The increase of the optimum bonding distance can be understood by the larger number of electronic shells of the host atoms.

The observed chemical y-offset for the 4th row elements (see figure 5.12) also appears in the case of the 5th and 6th row for filled d-shell host elements. It is based on the same chemical mechanism as discussed in section 5.3. The charge density difference is just slightly decreased by increasing the rows of the elements (see appendix). Looking at the chemical shifts in figure B.1, one might consider Hg to be an outlier, which does not fulfill the universal volume trend. For this case we note that the Hg look much more consistent with the other data, when studying the trends in the slope in B.2.

Detailed discussions of the physical mechanisms responsible for the chemical and mechanical part of the solution enthalpy would be very similar to those in 5.3. Instead of such a repetition, we continue our strategy to extend the database of solution enthalpies investigated in section 5.3 and put the data in a broader context by considering more interstitial elements.

5.5 Study of interstitial elements in transition metals

Beyond H, the interstitial elements considered in this work are He, B, C, N, O, F, and Ne. We next want to understand, if the trends discovered for H can also be transferred to these other interstitial atoms by following the philosophy of section 5.2 that strong correlations between the interstitial atoms implies similar physical mechanisms.

5.5.1 Chemical potential of the interstitial elements

The reference chemical potentials of the interstitial elements under focus, namely H, He, B, C, N, O, F, and Ne are determined. As mentioned before interstitial elements are influencing material properties through several mechanisms. According to Sievert's law the solution enthalpy is the key parameter which mainly determines the concentration of the element in the host matrix. The solution enthalpy ΔH is defined as follows

$$\Delta H = E_{mi} - E_m - \mu_i, \quad (5.8)$$

in which E_{mi} denotes the interstitial-metal system, E_m the pure metal system, and μ_i the chemical potential of the interstitial element.

A comparison of interstitial atoms requires a clear definition of their chemical potentials entering equation 3.44. As already mentioned in section 3.3 the chemical potential of the interstitial element depends on the environment. In this thesis, we assume an interaction with the gas phase. Therefore, a diatomic molecule in vacuum is selected in the case of the interstitial elements H, N, O, and F. Further, for the noble gases He and Ne a single atom is assumed. For the elements B and C, on the other hand, a diamond cell has been used as reference. The calculated chemical potentials and the bond lengths are shown in Table 5.7. The bond lengths can be compared with experiment and show a very good agreement with experiments. An overall good agreement for the calculated bond length of the diatomic molecules (lattice constant for the diamonds, respectively) is observed. The reliability of the determined chemical potentials is less clear, since this is a number that depends on the choice of the pseudopotential and cannot be measured directly. A very high accuracy is,

element	bond length (Å)		chem pot μ (eV/atom)
	gga-pbe	exp.	
single atom			
He	—	—	0.016
Ne	—	—	0.005
diatomic molecule			
H ₂	0.75	0.74	-3.375
N ₂	1.12	1.1	-8.3
O ₂	1.23	1.21	-4.38
F ₂	1.42	1.42	-1.85
diamond crystal			
B _{diam}	4.05		-5.733
C _{diam}	3.57	3.57	-9.094

Table 5.7: Reference chemical potentials of the studied interstitial elements. For the noble gas elements He and Ne it is a single atom in vacuum. The elements H, N, O, and F are calculated as diatomic molecules in vacuum. The interstitial elements B and C are assumed to be in the diamond structure.

however, not required, since the chemical potential only yields constant shifts on the trends for the solution enthalpy.

5.5.2 Interstitial solution enthalpy in NM fcc structure - d-band trend

In order to discuss the chemical trends, we have plotted the solution enthalpies of all interstitial elements considered here again as a function of position in the PTE, i.e. as a function of d-band filling. Plot 5.20 is therefore an extension of the figures 5.8 and 5.18 for the o-site, using again the constraint of non-magnetic fcc-structures. Due to this constraint we are not plotting the comparison with experimental data.

Similar to hydrogen, we also find for the other interstitial elements a clear dependence of the solution enthalpy on the filling of the d-shells. At the same time there is only little modification of this trend, when changing the element rows (from 4th to 6th row). We realize, however, that the actually shape of the d-band trend can strongly differ from one interstitial element to the next.

According to the ΔH trend in dependence of the d-band filling a rough separation of the interstitial elements into two groups can be made. The first group shows similar features as hydrogen and contains besides hydrogen the interstitial elements boron, carbon, and nitrogen. The trend is classified by a ΔH decrease in the beginning and then an increase of ΔH , while the late TMs (with filled d-shell) for the same row have almost the same solution enthalpy. The second group contains the interstitial elements helium, neon, and also fluorine. A characteristic behavior of this group is an increase of ΔH until the d-bands are half filled and then a continuous decrease of ΔH , including the late TMs. The interstitial element oxygen shows properties of both groups.

It is a straight-forward assumption that for the noble gases dominating the second group, the

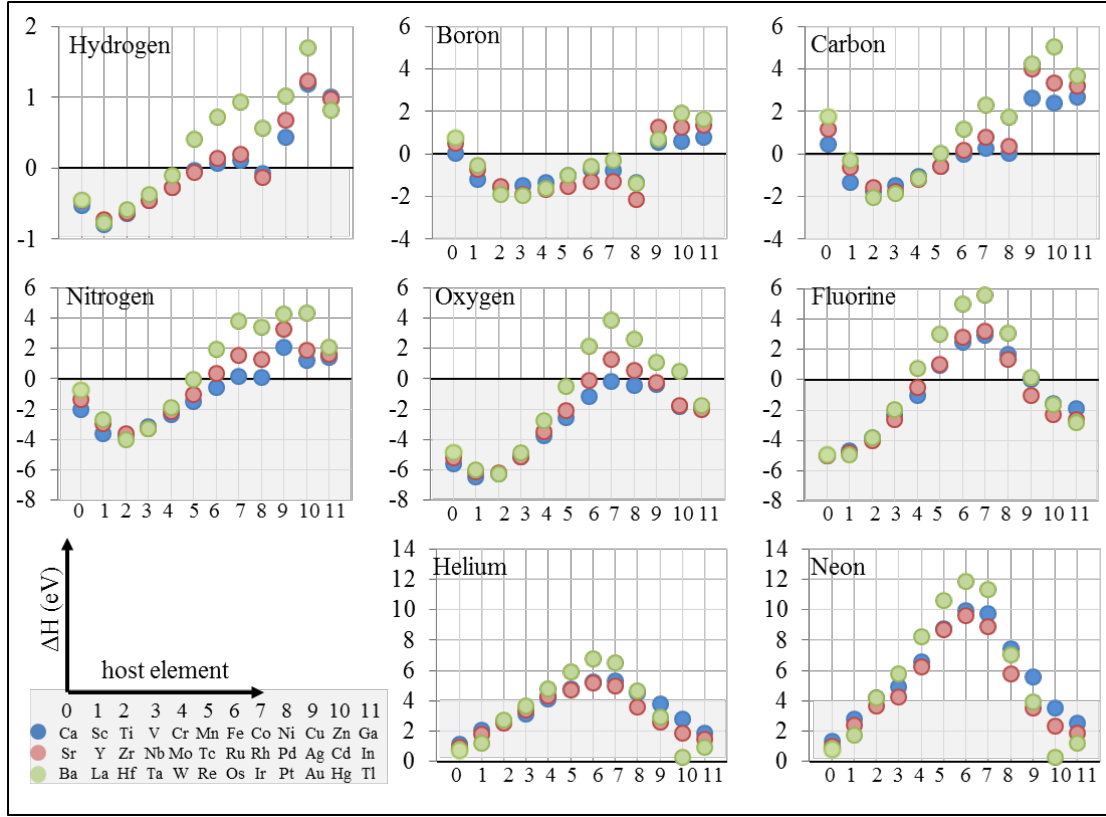


Figure 5.20: Solution enthalpy of interstitial elements in 4th, 5th, and 6th row elements in non-magnetic fcc structure.

chemical component of the solution enthalpy is less pronounced. This becomes in particular clear when studying again the universal trends as a function lattice constant. In figure 5.21 only the results for the octahedral site and host materials from the 4th row of the PTE are shown. A complete set of graphs for all studied host materials and for the octa- as well as tetrahedral sites can be found in section B of the appendix, which is similar to those plotted in figure 5.21. In the case of nitrogen and oxygen a similar chemical y-offset for a subgroup of host materials (mainly the late TMs) is observed as it was earlier observed for hydrogen (compare also figure 5.9). In the case of carbon and in particular boron the chemical y-offset for the individual host materials are so pronounced that even the impression of a universal solution enthalpy curve is destroyed (though the slope for a fixed host material still follows such a trend). The situation is completely different for the noble gases helium and neon. In these cases no shift between the intermediate and late TMs can be identified. In the case of fluorine this shift is very small. It is further remarkable that for this second group of elements the solution enthalpy in the octa- and tetrahedral site of the host material shows almost the same quantitative behavior (see figures in section B of the appendix), whereas noticeable difference in the shape and the absolute values are present for the first group.

These results confirm the assumption that the solution enthalpy (at least) for helium and neon must be dominated by an elastic effect. Indeed, looking at figure 5.20 we observe that the ΔH trend

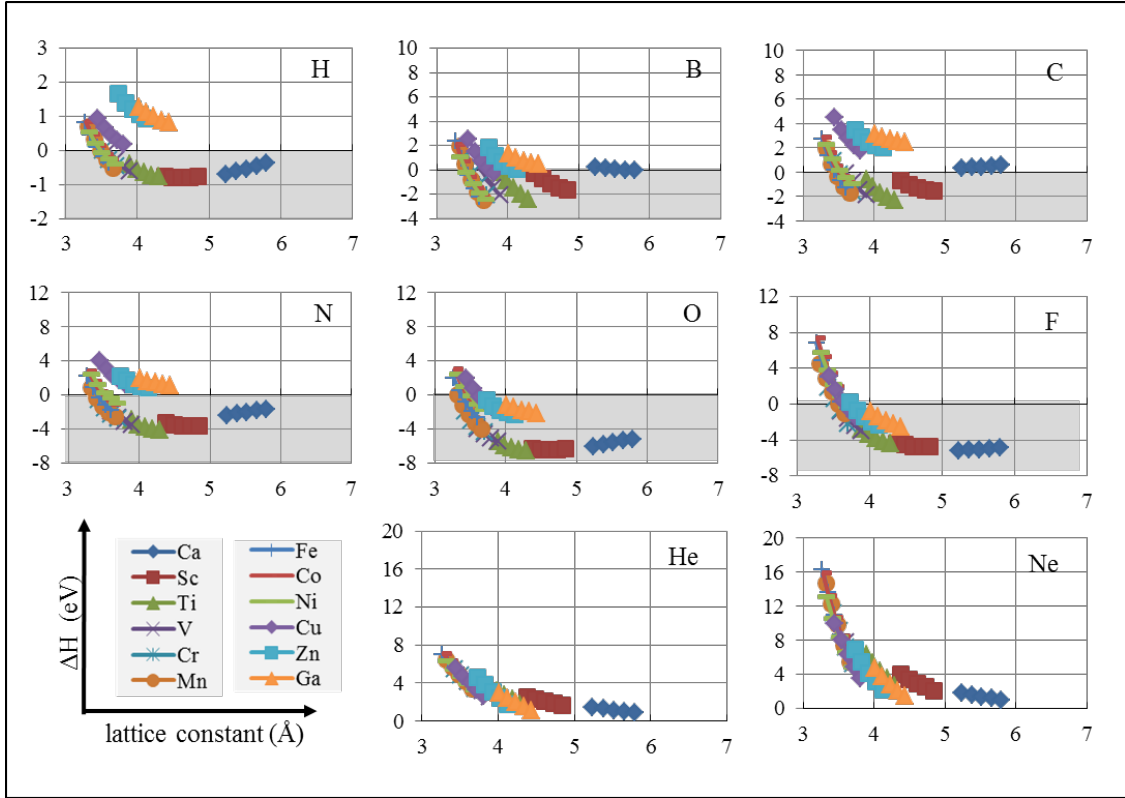


Figure 5.21: Solution enthalpy of interstitial elements in the octahedral site of 4th row host materials assuming a non-magnetic fcc structure.

for these interstitial elements is similar to that of the bulk modulus in figure 5.5b. A more quantitative analysis can be achieved by using statistical concepts as performed in the next subsection.

5.5.3 Interstitial ΔH vs B_0 and ΔE_{F-s}

The Bravais-Pearson correlation coefficient, which was introduced in section 3.6, is in this thesis used to optimize the correlation between the solution enthalpy and yet unknown (combinations of) host properties. In this section, the choice of the host properties is motivated by the observation of the previous section that depending on the interstitial element a mechanical or a chemical contribution might be dominating. The bulk modulus has been identified as a good descriptor of the former, whereas for the electronic contribution we follow the empirical considerations of Griessen (section 3.5) and choose the energy difference of the Fermi level and the center of the lowest lying conduction band ΔE_{F-s} .

The investigations are performed for the individual ground states (fcc, hcp, bcc, NM, FM, AFM) and the constraint to a NM fcc state, to improve the comparability. Considering the structure effect on the hydrogen solubility shown in table 5.6, modifications in the ordering of the elements when switching from one case to the other (top and bottom in figure 5.22) are not surprising. While central findings are not affected by these constraints of the structure, we provide precise numbers

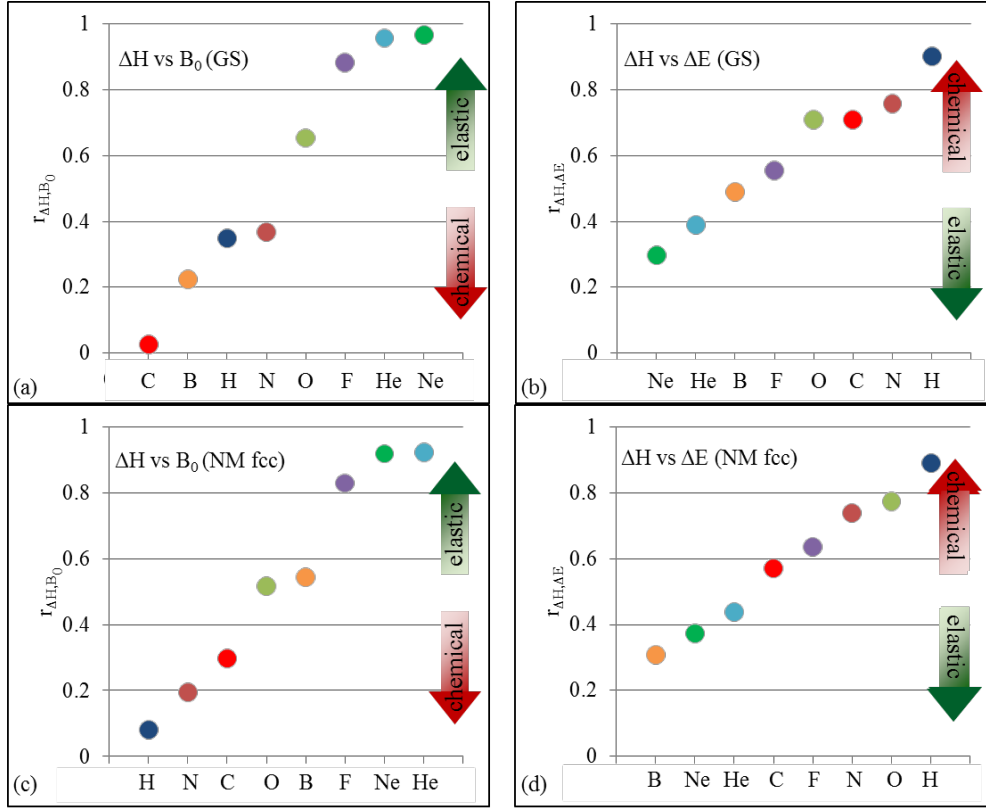


Figure 5.22: Correlation coefficients of the solution enthalpy ΔH for the interstitial element and the bulk modulus B_0 (left) or Griessen's ΔE_{F-s} (right) of the host elements in the ground state (top) or in the NM fcc state (bottom). The results are ordered according to the size of the correlation coefficient, indicating that noble gases correlate better with the bulk modulus (elastic), while H, O, C and N correlate better with ΔE_{F-s} (chemical). ΔH of boron doesn't correlate with both parameters.

mainly for the ground state (GS).

In figure 5.22 one can clearly identify the strong correlation of the solution enthalpy to the bulk modulus for the elements F, He and Ne (r_{xy} -values well above 0.8), confirming the previously assumed predominant mechanical nature of the solution process. In the case of He and Ne this correlation is with $r_{xy}^{GS} = 0.96, 0.97$ respectively, particularly high, which needs to be compared with the number for C ($r_{xy}^{GS} = 0.03$) on the lower side. Also F shows with $r_{xy}^{GS} = 0.88$ an unexpectedly strong elastic effect. On the other hand, the elements H ($r_{xy}^{GS} = 0.90$), N (0.76), O (0.71), and C (0.71) show a much stronger correlation with ΔE_{F-s} , though the r_{xy} -values are (apart from hydrogen) lower than 0.8. These results confirm the importance of the chemical contribution for these elements. Boron is special in the sense that it has for both properties correlation values ($r_{xy}^{GS} = 0.22$ and 0.49) below 0.6, i.e. it shows only weak correlations to a purely elastic or purely electronic description.

The high correlation values in the case of B_0 for the noble gases is reflected by a proportionality of ΔH and B_0 , i.e. an almost linear relation in the diagram. For the other group of interstitial

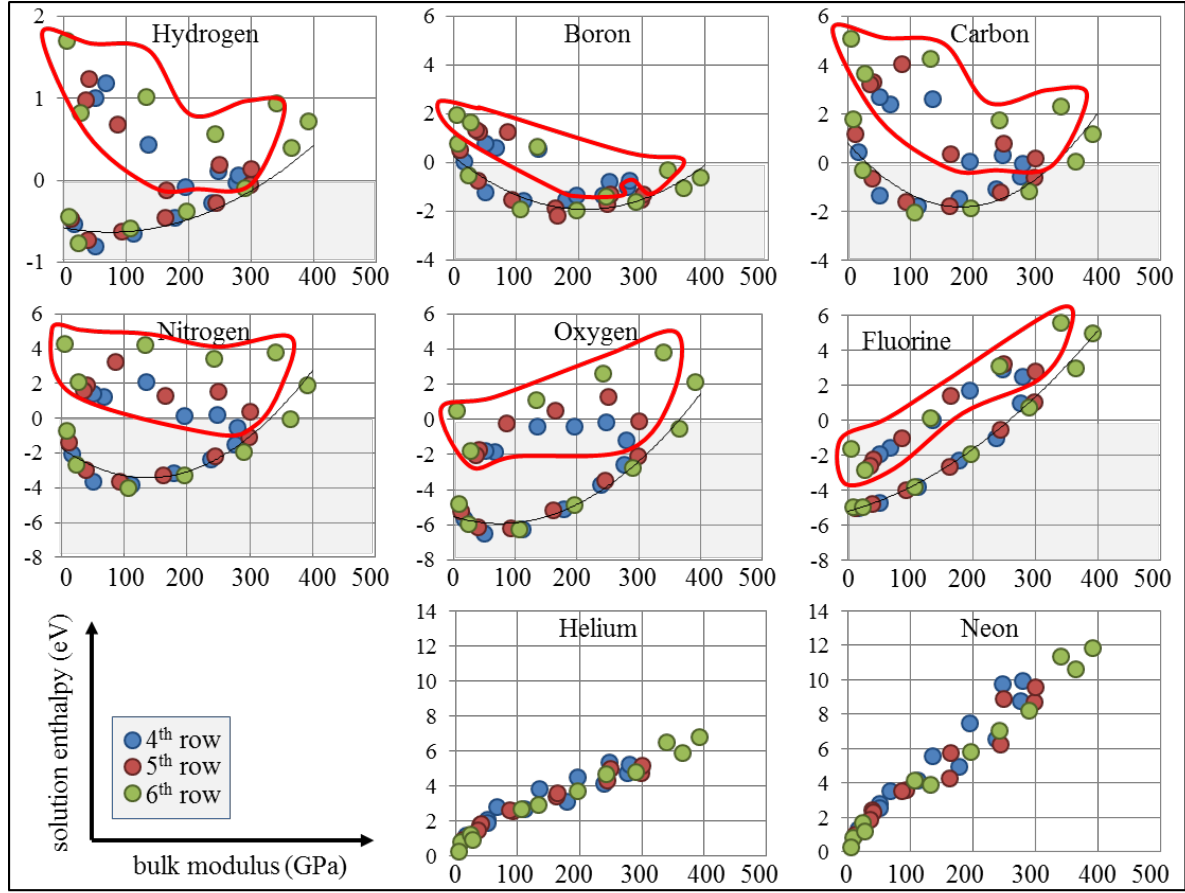


Figure 5.23: Solution enthalpy vs bulk modulus of interstitial elements in 4th, 5th, and 6th row elements in non-magnetic fcc structure. Elements with more than half-filled d-bands are red encircled. For below half-filled d-band matrix elements, the solution enthalpy shows a higher correlation to the bulk modulus.

elements (H, O, C, N) the data points appear to be much more scattered. To analyze this further, we first looked at F, an interstitial element that has also a rather large r_{xy} value. We realized that the dependence of the solution enthalpy on the bulk modulus separates into two branches. It turned out that all data points on the lower branch belong to host elements with less than half-filled outer d-bands, while the data points of the upper branch belong to host elements with more than half-filling.

Having identified this, one realizes that the lower branch is present for all interstitial elements. The slope of the solution enthalpy vs bulk modulus line correlates with the size of the interstitial atom. The larger the interstitial atomic number the larger the slope of the solution enthalpy line (slope(atomic number): He=0.0145(2), F=0.0227(9), Ne=0.028(10)). The data points for more than half-filled d-bands (red encircled in figure 5.23) can scatter much more heavily for those interstitial elements, where the elastic effects are insufficient to describe the solution enthalpy. An important outcome of this investigation is, therefore, that chemical trends can become much more pronounced

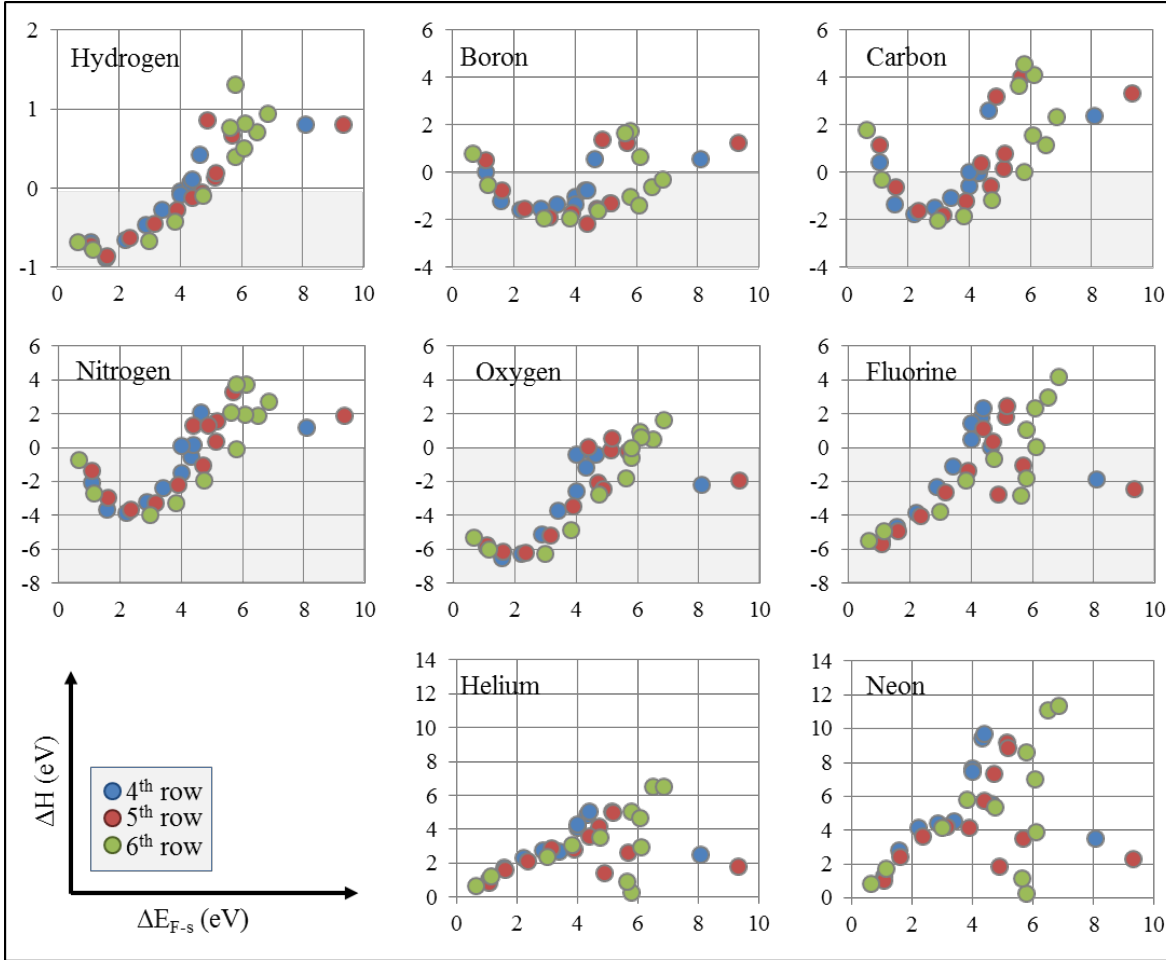


Figure 5.24: Solution enthalpy vs Griessen's ΔE_{F-s} of interstitial elements in 4th, 5th, and 6th row elements in non-magnetic fcc structure. A high correlation coefficient is achieved for hydrogen.

if only a part of the transition metals is taken into account.

5.5.4 Electronic effects and charge density differences

We have performed a similar analysis for the correlation of ΔH and the energy difference ΔE_{F-s} in figure 5.24. The good performance of hydrogen in this comparison is again clearly visible by a very good proportionality. The behavior of the other interstitial elements is less systematic and did not provide additional insights.

The chemical effect of interstitial solution is, however, multi-fold. On the one hand an interstitial atom will create energetically low lying states in the host matrix (bonding). The energy distance of the low-lying conduction bands to the Fermi energy is captured by ΔE_{F-s} . On the other hand additional electrons coming with the interstitial atom will occupy states at the Fermi-level. Since at the same time the filling of d-band shifts the d-band center to the left (negative energy w.r.t. E-fermi) a weakening of the interstitial-metal bonding can be expected. The latter kind of electronic

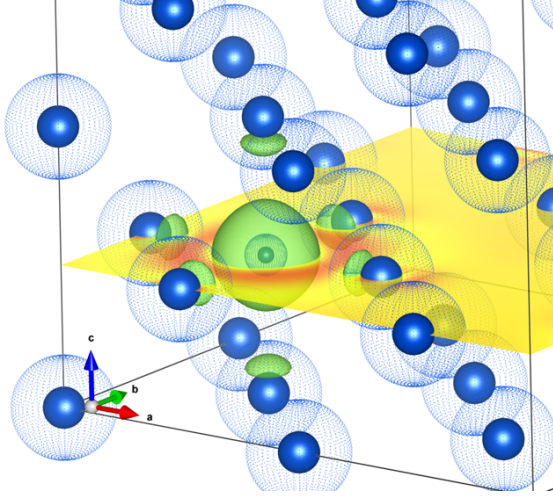


Figure 5.25: Example of the charge density difference induced by a H atom located in the octahedral site of NM fcc Ca. Green spherical isosurface represents the accumulation of electronic charge at the hydrogen atom. An isotropic charge accumulation around the hydrogen atom is for all elements observed.

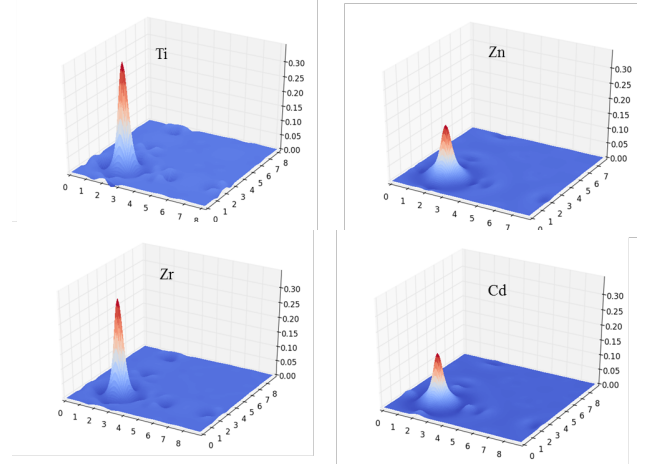


Figure 5.26: Charge density difference for H in early transition metals (Ti and Zr) compared to late transition metals (Zn and Cd). The peaks show the charge accumulation at hydrogen position. An increased density for hydrogen soluble elements is observed.

effects can be captured by charge-density differences (CDDs).

Indeed, we have shown in figure 5.16 for the example of hydrogen that the spherically integrated CDD is related to the chemical component in the solution enthalpy as well as to the element-specific chemical shift in the universal curves. The CCD of hydrogen in the four host materials Ti, Zn, Zr, and Cd is shown in figure 5.26. It compares two early transition metals (Ti and Zr) to two late transition metals (Zn and Cd), using this as an explanation for the observed chemical shifts. We therefore tried to generalize this approach to the other interstitial elements.

Charge density difference (H in NM fcc TM)											
Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga
0.27	0.32	0.35	0.34	0.34	0.31	0.29	0.25	0.21	0.21	0.17	0.22
Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In
0.26	0.30	0.32	0.30	0.27	0.23	0.19	0.14	0.08	0.14	0.16	0.22
Ba	La	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Tl
0.27	0.32	0.31	0.27	0.25	0.19	0.14	0.08	0.01	0.02	0.07	0.21

Table 5.8: Charge density difference of H in TM. A charge accumulation at the H position is observed for all studied elements. The table shows the maximum (peak) value of figure 5.26

However, the simple picture of hydrogen bonding is not applicable to the other interstitial elements since they have 2p band electrons, which make the chemical behavior more complex. In figure 5.27 the interstitial elements CDD in the host element Fe is shown. A more complete set of CDD figures is provided in B.4. The p-bondings character can be seen, but not a peak in the center of the interstitial element like in the case of H. The peak in the case of neon occurs due to the 'squeezing' of the neon electrons (see appendix). We conclude that the best justified parameter for the description of the chemical effect is still missing. We will next use statistical concepts to determine it.

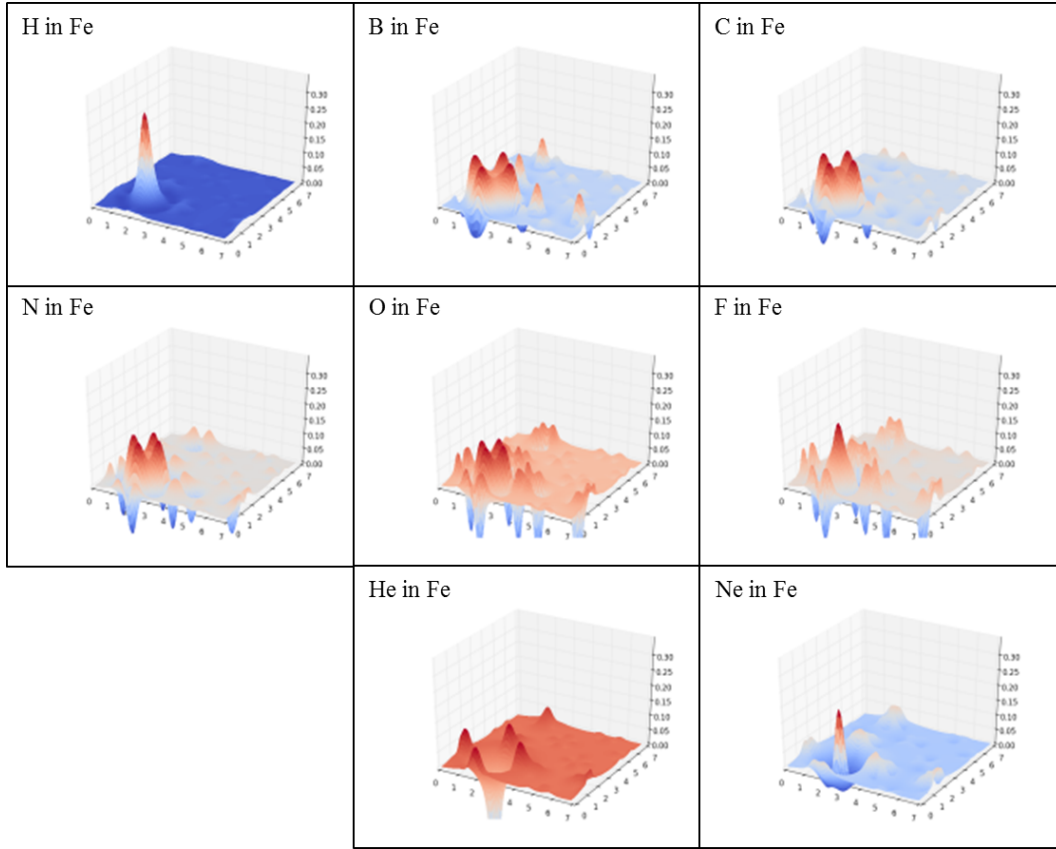


Figure 5.27: Charge density difference for H, B, C, N, O, F, He, and Ne in Fe. The single peak for hydrogen cannot be observed for the other interstitials.

Chapter 6

Results II: Knowledge Discovery in Databases

This chapter combines the goal of finding reliable descriptors for the prediction of the solubility of interstitial elements with high-throughput approaches outlined in chapter 2. In the previous chapter the empirical approach of Griessen for hydrogen solution enthalpies (section 3.5), has been used as a starting point for such an investigation. We realized, however, in section 5.5 already that the extension of the database to other interstitial elements yields new insights and candidates for descriptors such as the bulk modulus B_0 for the host elements assumed to be in non-magnetic fcc structure. In the present chapter, the analysis of our database shall be performed more systematically to identify the a priori unknown optimum parameter combination for the prediction of solution enthalpies for the host elements in the ground state.

6.1 Correlations between solution enthalpies of interstitial elements

We first exploit the observation that there are large correlation coefficients between the experimental solution enthalpies of hydrides and borides or between hydrides and nitrides as shown in section 5.2. We want to transfer this notion to the dilute limit considered in the present thesis. High correlation coefficients between the solution enthalpies of the interstitial elements H, B, C, N, O, F, He, and Ne can be expected, if they depend on the same physical properties of the pure host material Γ_y , i.e. parameters that are independent of the interstitial. This is captured by the following predictive formula for the solution enthalpy of an interstitial element x in a TM y

$$\Delta H_{xy} \approx a_x \Gamma_y + b_x, \quad (6.1)$$

in which a_x and b_x denote now fitting parameters that only depend on the interstitial. The previously discussed equation 3.74 of Griessen has such a form. It contains the d-band width W_d , the distance between the interstitial element and the nearest neighboring host atoms $\sum_j R_j$ (the sum denotes the coordination number of the interstitial element, i.e. six for octahedral site, four for tetrahedral site respectively), the energy difference of the Fermi level and the energy of the lowest lying s-like

conduction band $\Delta E_{F-s} = E_F - E_s$, combined as

$$\Gamma_y = \Delta E_{F-s} \cdot W_d^{1/2} \cdot \sum_j R_j^{-4}. \quad (6.2)$$

However, it is not clear whether this is the most suitable choice for hydrogen and in particular for the whole set of interstitial elements.

6.1.1 Linear dependence between solution enthalpies

The solution enthalpies of individual interstitial hydrogen, boron, carbon, nitrogen, oxygen, fluorine, helium, and neon atoms (i.e. their dilute limit) in 3d, 4d, and 5d transition metals (TM) are determined. The host materials are considered in their structural and magnetic ground state (GS), where we restrict ourselves to those transition metals that are FCC, BCC, or HCP (NM or FM) in their ground state. Elements with a more complicated structural or magnetic ground state (e.g., Mn and Cr) as well as non-transition metals (e.g., Ca and Zn) are not taken into account. The relevant interstitial positions are the octahedral or the tetrahedral site, of which the energetically more favorable site determines the solution enthalpy ΔH .

If the same bulk properties Γ_y of the host material y determine the solution enthalpy of two different interstitial elements (for example hydrogen (ΔH_{Hy}) and carbon (ΔH_{Cy})) then equation 6.1 can be applied twice and the solution enthalpy of, for example, carbon in the host material y can be determined through the equation

$$\Delta H_{Cy} = \frac{a_C}{a_H} \Delta H_{Hy} - \frac{a_C b_H}{a_H} + b_C, \quad (6.3)$$

in which the subindices H and C corresponds to the interstitial elements hydrogen and carbon and the coefficients a and b are the fitting parameters in equation 6.1. Therefore, the approach yields a linear dependence between both enthalpies. A quantitative analysis for the linearity is given by the Pearson correlation coefficient (PCC), which is close to 1 for perfect correlation and 0 if no correlation is present. The results of the solution enthalpies for the interstitial elements H, B, C, N, O, F, He, and Ne in 3d (first block), 4d (second block) and 5d host elements (third block) is presented in table C.1.

The results in table C.1 allows the determination of the PCC matrix for all pairs of interstitial elements in the considered set of transition metals as given in table 6.1.

The PCC matrix elements for the ground state structures show a high correlation coefficient between solution enthalpies of interstitial elements that have neighboring positions in the PTE (figure 5.1). For example, it is $PCC_{C,N} = 0.94$ for the solution enthalpies of C and N and $PCC_{N,O} = 0.96$ for N and O. Such a behavior is not very surprising since these interstitial elements only differ by one missing or additional $2p$ electron in its shell. Less obvious are the high correlation coefficients between hydrogen and the elements carbon, nitrogen, and oxygen with 0.85, 0.88, and 0.86. This is interesting since the chemical contribution to the solution enthalpy of hydrogen has been earlier (see section 5.3.6) connected with the *anion model*, which might be specific to the electronic configuration of hydrogen with only one $1s$ electron. It is therefore remarkable that carbon, nitrogen, or oxygen show a chemically different behavior.

	H	B	C	N	O	F	He	Ne
H	1.00	0.65	0.85	0.88	0.86	0.8	0.71	0.59
B		1.00	0.75	0.56	0.42	0.31	0.25	0.21
C			1.00	0.94	0.83	0.64	0.41	0.28
N				1.00	0.96	0.83	0.6	0.47
O					1.00	0.94	0.76	0.65
F						1.00	0.92	0.86
He							1.00	0.97
Ne								1.00

Table 6.1: PCC matrix for all pairs of interstitial elements. The host elements are considered in their ground state structure and magnetic phase. Further, they are limited to transition metals, i.e. without the columns IIa, IIb and IIIa, which are also highlighted in figure 5.1.

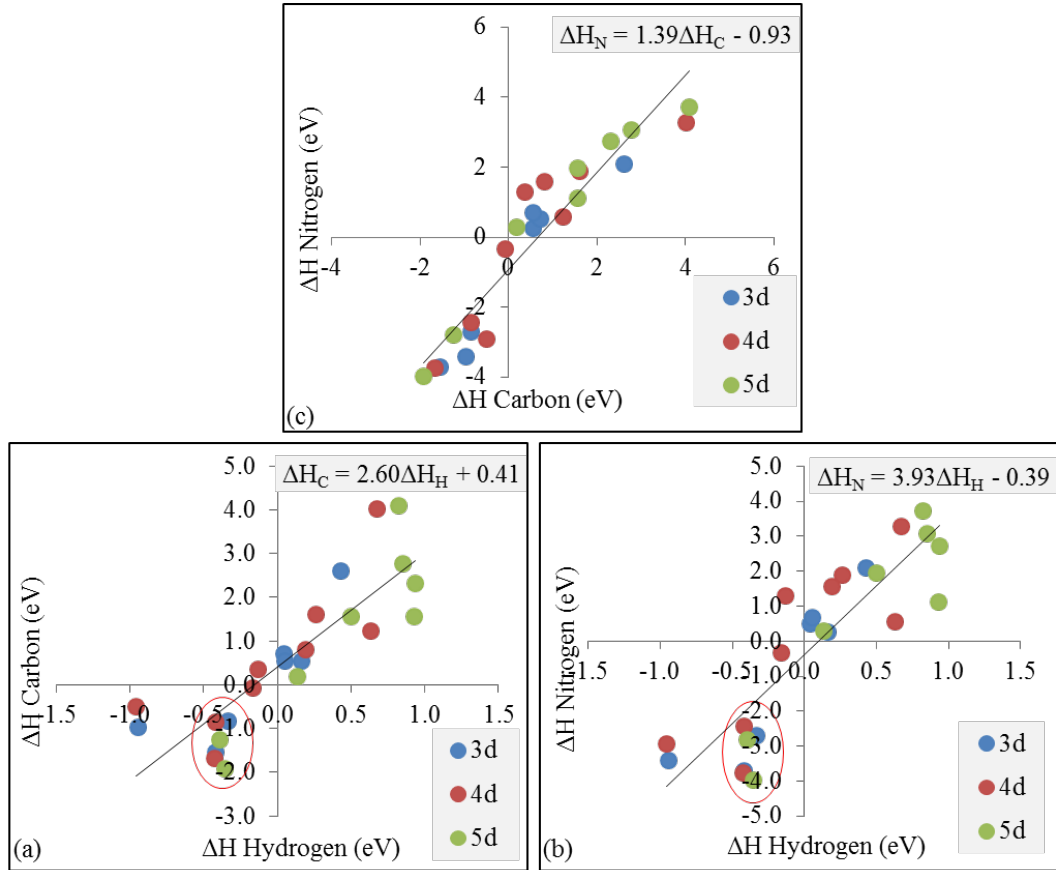


Figure 6.1: Linear dependence of solution enthalpies of interstitial elements in the dilute limit. The slope of the regression line in figure (a), which is the first quotient $\frac{a_{\text{C}}}{a_{\text{H}}}$ of equation 6.3, indicates a carbon-hydrogen solution enthalpy relation of approx 2.6, (b) a nitrogen-hydrogen relation of ≈ 3.9 and (c) a nitrogen-carbon relation of ≈ 1.4 .

A second interesting effect characterizes the first row of the PCC matrix (i.e. the PCC between the hydrogen atom and the $2p$ interstitial elements). Here, the highest correlation is reached with the interstitial element N, which has a $2p$ band that is exactly half filled with three electrons (same as hydrogen has a half filled $1s$ band containing one electron). A symmetric behavior is observed and to the left (clearing the $2p$ shell) and right (filling the $2p$ shell) of N the PCC is decreasing in the same way.

The correlation coefficients between the solution enthalpy of H and C or H and N (figure 6.1) show a clear trend: with increasing hydrogen solubility the N (or C) solubility is increasing with a certain factor (given by the slope of the solid line) and vice versa. This trend is not perfectly fulfilled for all host elements. Red circles in figure 6.1 mark some of these exceptions. Increasing C (or N) solubility doesn't mean an increased solubility for H for these host elements. These are mainly the early (3d, 4d and 5d) transition metals. However, the effect disappears if the interstitial elements C and N are compared with each other which have a correlation coefficient of 0.94.

Interesting is also the size of the slope. A slope of 3.93 in the H-N curve (figure 6.1(b)) means: if a change of the host material increases the H solution enthalpy by 1 eV, then the same change yields an increase of the N solution enthalpy by almost 4 eV. These slopes in the correlations between H, N, and C can be directly connected with the electronic configuration of the interstitial elements.

6.1.2 Chemical vs elastic effect on the interstitial elements

The predictive formula 6.1 assumes for the interstitial elements with a high correlation coefficient (for example H and N) physical properties Γ_y of the host materials, which influence their solubility in the same way. The decisive question is whether the physical properties we discussed in chapter 5 in the context of hydrogen provide the most reliable descriptors. As indicated there, the solution enthalpy should be separated into an elastic term (relaxation effect) and a chemical term (bonding/anti-bonding). We start with the elastic term and revisit the bulk modulus, which has been identified in section 5.5 as a useful parameter for describing the elasticity of the host material. As already discussed in section 5.5.3 the noble gases (He and Ne) as well as F show a high correlation between their solution enthalpy and the bulk modulus (i.e. the elastic effect dominates).

The bulk modulus of the host material is one candidate to be used for the unknown parameter Γ_y directly. We should mention that B_0 is not the only possible choice for Γ_y , but also (a combination of) physical parameters which are not considered yet could be possible.

Helium vs Neon

Noble gases such as helium and neon do not show a significant chemical effect influencing its solution enthalpy in the TMs. Instead, the solution enthalpy is mainly influenced by an elastic effect (relaxation effect of the neighboring host atoms). The solution enthalpy of the interstitial elements He and Ne is plotted with respect to the bulk modulus of (i.e. representative for the elasticity of) the host material. This is shown in figure 6.2(a).

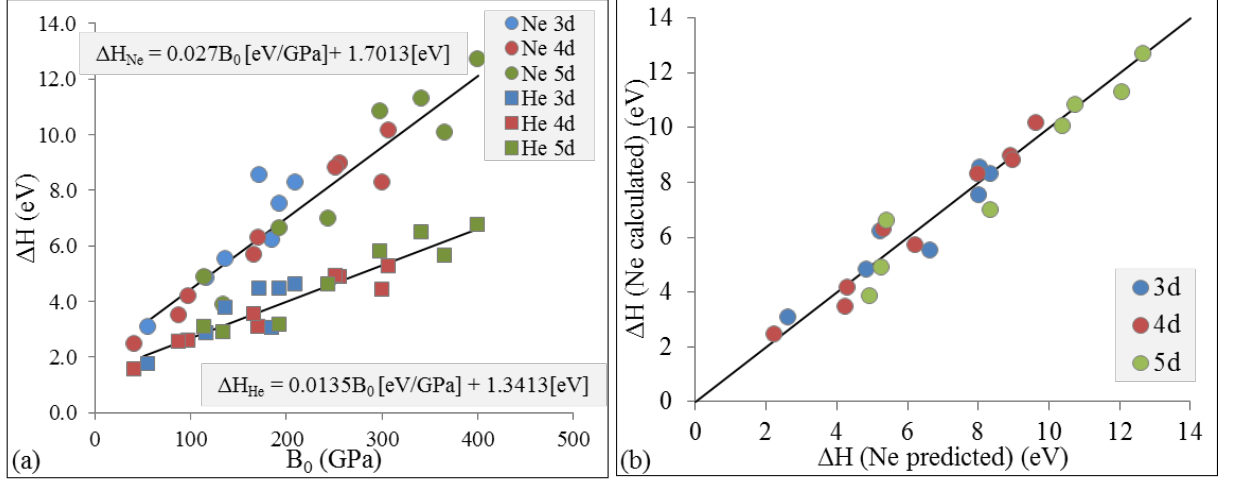


Figure 6.2: (a) Solution enthalpy vs bulk modulus of He (blue) and Ne (red) in 3d, 4d, and 5d TMs in the ground state. The interstitial element neon is approximately two times less soluble than helium in the same material. (b) Comparison between the predicted solution enthalpy of Ne by using equation 6.4 and the calculated solution enthalpy of Ne in the same host material.

A regression line for the data points in figure 6.2(a) provides the interstitial-dependent terms of equation 6.1 as $a_{He} = 0.0135$ and $b_{He} = 1.3413$ for helium, and $a_{Ne} = 0.027$ and $b_{Ne} = 1.7013$ for neon. The slope of Ne is, therefore, exactly two times larger than the slope of helium. In other words, the solubility of the larger interstitial element neon is decreasing two times faster with increasing bulk modulus than that of helium. This is in remarkable agreement with the ratio of covalent radii of both elements ($Ne/He = 0.58/0.28 = 2.07$). The correlation 6.3 of both interstitial elements is therefore of the form

$$\Delta H_{Ne,y} = \frac{a_{Ne}}{a_{He}} \Delta H_{He,y} - \frac{a_{Ne}}{a_{He}} b_{He} + b_{Ne} = 2\Delta H_{He,y} - 0.98, \quad (6.4)$$

which can be used to predict the solution enthalpy of Ne in a host element y if the solution enthalpy of He in the same host element is known. The predictive power of this approach is shown in the figure 6.2(b).

A similar predictive formula can be obtained for all other interstitial elements. If the solution enthalpies are correlated it can then be reduced to a lower number of variables with the factor analysis, which was introduced in section 3.7 and will be applied in the following.

6.1.3 Sharpening chemical trends

We realized in section 5.5.3 that chemical trends and materials laws can become significantly more pronounced, if the set of considered host materials is reduced. In the case presented there (see figure 5.23) the descriptor was already identified to be the bulk modulus. For the more general approach considered here, the correlation matrix (table 6.1) is used as a criterion.

It is obvious that the reduction of the database (i.e. the reduction of the number of host materials) will increase the coefficients in the correlation matrix. One can iteratively investigate

for which of the host materials this effect is most significant and remove it. It turns out that this procedure would result in a removal in the order Ag, Au, Pd, Cu, Pt, Rh, Y, Re, Ni, Tc, Sc, Ir, Ru, Co, Fe, and Ta. The impact on the correlation matrix (table 6.1) is shown in figure 6.3. The left figure shows the average value of the whole correlation matrix r_{avg} and the right figure shows the average value of the rows (i.e. the impact on a certain interstitial element). It turns out that the first six elements (from Ag to Rh) which are eliminated are all (six of eight) fcc elements in ground state, and the first four eliminated elements have filled d-shells. By removing the host elements Ag and Au the average value of the correlation matrix is raising above 0.8. Analyzing the row average value of the correlation matrix yields the result that the interstitial element boron shows the lowest correlation to the other interstitial elements. Even if one would remove six host elements, the average value of the boron row is still around 0.7 and thus below the other interstitial elements (all around 0.9).

We conclude from these observations that the goal of finding a universal predictive equation for the interstitial elements solubility is easier achieved if the interstitial element boron and the host elements Ag and Au are removed from the studied data.

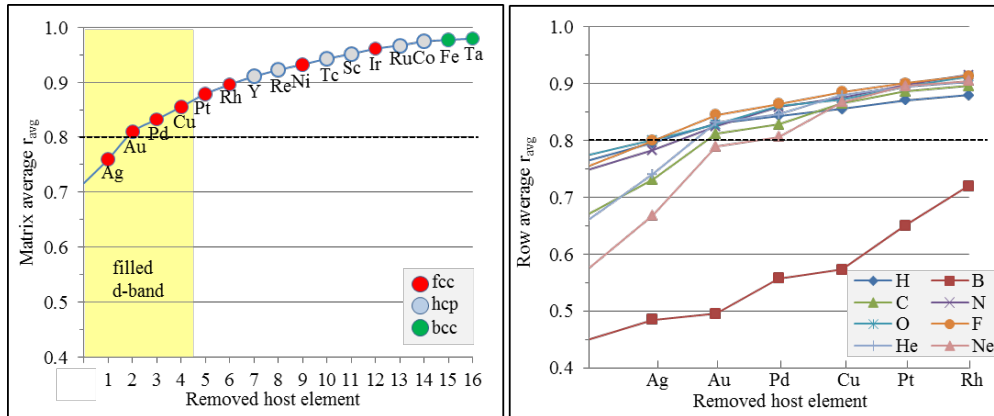


Figure 6.3: Left: Average value of the changed correlation matrix 6.1 after reducing the host elements with the most adverse effect on the correlation. Right: Average value of the rows of the same matrix. It turns out that fcc elements (or the filled d-shell elements) have the lowest impact on the correlation coefficients. The first six elements are fcc structure in GS and the first four of them have filled d-shells.

6.2 Factor analysis of the solution enthalpies

The factor analysis (FA) has been introduced in section 3.7 as a method to reduce a system with m observed variables to a system with ($k < m$) latent variables, which are called factors. It is applied in this section to the matrix given in table 6.1 that summarizes the correlation coefficients between the solution enthalpies of the interstitial elements H, B, C, N, O, F, He, and Ne in transition metals. Details of this analysis are given in appendix C, while here the main outcomes are provided.

6.2.1 Determination of the number of factors k

In order to determine the number of required factors k , a singular value decomposition needs to be performed first. This yields the eigenvalues (5.3876, 1.4730, 0.5436, 0.1184, 0.0170, 0.0071, 0.0025, 0.0008) of our initial matrix 6.1. According to the Kaiser-Guttman criterion, k is then given by the number of eigenvalues that are larger than 1. As a consequence we have worked with a reduction to two factors, for which we don't know the physical meaning at this stage.

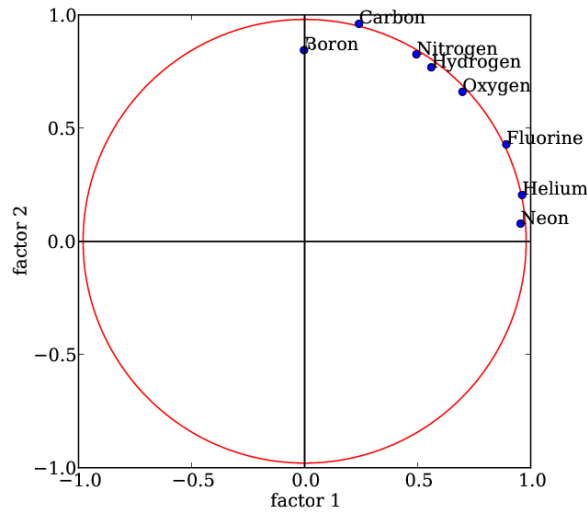


Figure 6.4: Correlation circle resulting from a principal component analysis after the reduction to two factors and an anti-clockwise rotation by $\alpha = 43.92^\circ$. Points on the circle with radius 1 are well described by the two factors. Interstitial elements that appear next to each other close to the circle are significantly correlated. Boron is an outlier in this respect.

The quality of this reduction is best visualized by a correlation cycle resulting from the principle component analysis as given in figure 6.4. It projects the full information of the correlation matrix on the two factors. The red circle has a radius 1 and is the set of all points for which this reduction is sufficient. This seems to be the case for all interstitial elements except boron. Points next to the circle, which are close to each other are strongly correlated. This applies mainly to the interstitial elements that are neighbors in the PTE. The representation confirms that hydrogen and nitrogen are strongly correlated (r close to 1), as well as helium and neon. Points next to the circle that differ by an angle of 90° , are uncorrelated (r close to 0). This is for example the case for carbon and neon.

Boron seems to be an outlier, consistent with our observation in section 6.1.3 that it shows the lowest correlation with the other interstitial elements. In the present context this means that the two factors identified for the other interstitial elements do not completely capture the physics of boron solubility. The difference of the full correlation matrix 6.1 and the reduced correlation matrix

$\mathbf{R}_{\text{red}^2}$,

$$\mathbf{R}_{\text{red}^2} - \mathbf{R} = \begin{pmatrix} -0.10 & -0.01 & 0.02 & 0.04 & 0.03 & 0.03 & -0.02 & 0.00 \\ -0.01 & -0.29 & 0.06 & 0.13 & 0.13 & 0.05 & -0.08 & -0.15 \\ 0.02 & 0.06 & -0.02 & -0.03 & -0.03 & -0.01 & 0.01 & 0.03 \\ 0.04 & 0.13 & -0.03 & -0.07 & -0.07 & -0.03 & 0.04 & 0.07 \\ 0.03 & 0.13 & -0.03 & -0.07 & -0.08 & -0.03 & 0.04 & 0.07 \\ 0.03 & 0.05 & -0.01 & -0.03 & -0.03 & -0.02 & 0.02 & 0.03 \\ -0.02 & -0.08 & 0.01 & 0.04 & 0.04 & 0.02 & -0.03 & -0.04 \\ 0.00 & -0.15 & 0.03 & 0.07 & 0.07 & 0.03 & -0.04 & -0.08 \end{pmatrix}, \quad (6.5)$$

has by far the largest number in the boron column (up 0.29) compared to any other column (all below 0.1). Consequently, an exclusion of boron would result in a plot like figure 6.4 in which all elements are almost lying on the circle with radius one. If instead of two a set of three factors is used, then the correlation coefficients are for boron, but also for all other elements much better captured. This is again documented by the difference between the original PCC and the three-components reduced PCC:

$$\mathbf{R}_{\text{red}^3} - \mathbf{R} = \begin{pmatrix} -0.09 & 0.03 & 0.01 & 0.02 & 0.02 & 0.02 & -0.01 & 0.02 \\ 0.03 & -0.01 & 0.00 & -0.01 & -0.01 & -0.01 & 0.00 & -0.01 \\ 0.01 & 0.00 & -0.01 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.02 & -0.01 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.02 & -0.01 & 0.00 & 0.00 & -0.01 & 0.00 & 0.00 & 0.00 \\ 0.02 & -0.01 & 0.00 & 0.00 & 0.00 & -0.01 & 0.01 & 0.00 \\ -0.01 & 0.00 & 0.00 & 0.00 & 0.00 & 0.01 & -0.01 & 0.01 \\ 0.02 & -0.01 & 0.00 & 0.00 & 0.00 & 0.00 & 0.01 & -0.01 \end{pmatrix}. \quad (6.6)$$

One can clearly see that the reduction to three factors (instead of eight) is completely sufficient to describe the similarity of the solution enthalpy mechanism between all interstitial elements. Therefore, we will use for the predictive formula search up to three addends (equal to three common factors), each of which can be a product of maximum three pure bulk properties (which is an arbitrary choice in order to limit the number of possible combinations).

The approach needs to be compared to the empirical ansatz 3.74 for hydrogen only, where two addends (one constant and one factor containing a product of exact three bulk properties 3.70) have been used.

The factor analysis with three factors (figures 6.5 and 6.6) reveals that the interstitial element boron has a dominant third factor (labeled as 3), which is less dominant for the elements hydrogen, carbon and nitrogen and not relevant for the remaining interstitial elements oxygen, fluorine, helium, and neon. In addition the accuracy of using three factors is documented by the fact that the radial parts of the interstitial elements are with $r \approx 0.995$ very close to one. Hydrogen has a slightly lower radial part with $r = 0.95$. Using the information obtained in the previous sections on the mechanical nature of the solution enthalpies of F, He and Ne, it can be assumed that factor 3 is mainly of chemical nature. A more detailed information about the physical properties described by the three factors is, however, not accessible by the factor analysis.

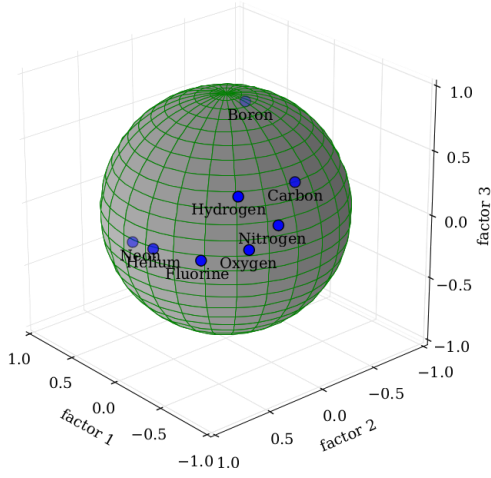


Figure 6.5: Result of the principal component analysis after the reduction to three factors.

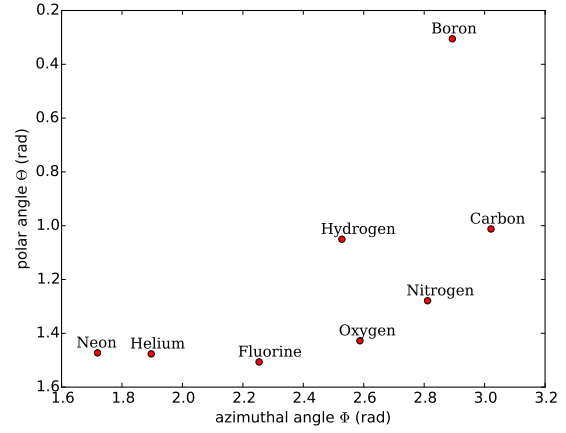


Figure 6.6: Angle view of the principal component analysis after the reduction to three factors. The radial parts are not plotted here since almost all are with $r = 0.99$ very close to one (except hydrogen with $r = 0.95$).

6.2.2 Factor analysis for NM fcc structure

In section 6.1.3 we have investigated, which of the considered host elements have the least systematic behavior for the interstitial solubilities and realized that this are primarily fcc metals. This raises the question how the fcc structure influences the correlation of ΔH . For this reason, the correlation matrix for the solution enthalpy of the interstitial atoms given in table 6.1 is again calculated for the constraint that the host elements are forced to be in a NM fcc structure. The result is shown in table 6.2.

It turns out that the differences between the correlation matrices for ground state host elements and the same elements in NM fcc structure are small. In the hydrogen row, for example, most of the deviations are in the order of 2 %. Significantly larger deviations are only obtained for the smallest absolute values, with a maximum deviation of 0.15 for the boron-neon correlation. We can therefore conclude that the structural differences of the host elements is not drastically influencing the correlation among the solution enthalpies of the interstitial elements. To keep the predictions as realistic as possible, we therefore perform the upcoming investigations for the respective ground state of the host materials.

6.2.3 Search for the best ΔH predicting equation and corresponding bulk parameters

Based on the factor analysis we can now approach the problem to find the most suitable descriptors in a predictive formula of the interstitial solubility similar to what has been established in equation 6.1. It is important to stress again that foreseen physical parameters Γ_y should solely describe the host system y , without taking the interstitial x into account. The list of candidate parameters is based on the experiences with ΔH of hydrogen as expressed in equation 3.74, where $\Gamma_y = \Delta E \cdot W_d \cdot \sum_j R_j$.

	H	B	C	N	O	F	He	Ne
H	1.00	0.65	0.87	0.93	0.92	0.82	0.73	0.61
B		1.00	0.85	0.67	0.47	0.23	0.15	0.06
C			1.00	0.94	0.79	0.56	0.4	0.27
N				1.00	0.94	0.77	0.62	0.5
O					1.00	0.93	0.8	0.71
F						1.00	0.94	0.89
He							1.00	0.98
Ne								1.00

Table 6.2: Pearson correlation matrix as in table 6.1, but for transition metals calculated in the NM fcc structure.

In addition our considerations in section 5.5 motivated us to consider the bulk modulus as descriptor for the mechanical properties. In order to take into account that the chemical trends for hydrogen (for example, the anion model was applicable for hydrogen) are not reproduced in the same way for all other interstitial elements, we also extend the spectrum of considered electronic properties. Both, the total density of states for the pure host element as well as its d-band projection are taken into account.

In summary, the list of considered candidates for descriptors include:

- the width of the d-band $W_d^2 = \frac{\int \rho E^2 dE}{\int \rho dE}$ (its 4th root was used by Griessen in his model 3.5),
- the interstitial distance to the nearest neighbor atoms $\sum_j R_j$ (different to Griessens approach 3.5 we sum over both interstitial sites and divide by two),
- the energy difference between the energy at the s-like state and the Fermi-energy ΔE_{F-s} (used by Griessen 3.5),
- the bulk modulus $B_0 = -V \frac{\partial p}{\partial V}$,
- the total number of states in the band structure of the solid $N_{tot} = \int_{-\infty}^{\infty} \rho dE$,
- the number of occupied states in the solid $N_{occ} = \int_{-\infty}^{E_F} \rho dE$,
- the fractional band filling $f_d = \frac{N_{occ}}{N_{tot}}$, and
- the average energy of the d-band (d-band center) $\epsilon_d = \frac{\int \rho E dE}{\int \rho dE}$.

The results for the bulk properties are given in table 6.3.

Besides the number of possible descriptors, we also extend the structure of the predictive formula. The version

$$a_x \cdot \Gamma_y + b_x \approx \Delta H_{xy}, \quad (6.7)$$

introduced in equation 6.1, describes the enthalpy and the site-occupation of hydrogen very well and is also fulfilling physical constraints. The formula is also working well for neighboring interstitial elements of the PTE, such as Ne and He in neighboring rows or F and Ne in neighboring

	$W_d^{1/2}$ (eV)	R_j (\AA^{-4})	ΔE_{F-s} (eV)	B_0 (GPa)	f_d	ϵ_d (eV)	N_{tot}	N_{occ}
Sc	1.69	0.23	1.54	54	0.13	2.28	7.83	0.99
Ti	1.59	0.37	2.34	116	0.23	1.51	8.33	1.93
V	1.65	0.57	2.64	184	0.35	1.35	8.88	3.10
Fe	1.52	0.69	3.00	170	0.66	-0.59	9.32	6.16
Co	1.54	0.69	3.50	208	0.77	-1.26	9.44	7.23
Ni	1.51	0.68	3.51	191	0.87	-1.55	9.48	8.25
Cu	1.71	0.60	4.66	136	0.97	-2.62	9.56	9.24
Y	1.86	0.16	1.55	40	0.15	2.68	7.72	1.15
Zr	1.81	0.25	2.49	96	0.25	1.98	8.03	1.99
Nb	1.92	0.37	2.86	169	0.32	2.11	7.00	2.27
Mo	1.77	0.45	4.42	255	0.45	0.17	8.59	3.89
Tc	1.75	0.47	4.61	299	0.56	-0.20	8.91	5.02
Ru	1.76	0.50	5.33	306	0.66	-1.31	8.95	5.94
Rh	1.74	0.48	5.18	250	0.77	-1.82	9.08	7.03
Pd	1.62	0.43	4.38	165	0.89	-1.78	9.22	8.18
Ag	2.08	0.35	5.70	87	0.98	-4.10	9.44	9.24
Hf	1.99	0.26	3.20	113	0.23	2.47	7.71	1.80
Ta	2.02	0.37	3.51	192	0.34	2.28	7.81	2.66
W	1.88	0.44	5.31	297	0.43	0.44	8.12	3.52
Re	1.87	0.45	5.56	365	0.53	0.05	8.41	4.50
Os	1.91	0.48	6.72	399	0.64	-1.47	8.58	5.48
Ir	1.93	0.46	6.84	341	0.75	-2.29	8.78	6.57
Pt	1.84	0.42	6.09	243	0.86	-2.41	8.96	7.74
Au	2.00	0.35	6.13	133	0.97	-3.62	9.16	8.86
Avg. (μ)	1.79	0.44	4.21	200	0.57	-0.32	8.64	5.11

Table 6.3: The bulk properties of the pure host elements. The corresponding average values and standard deviations of the properties (μ and σ) are shown below. The last line indicates the average values after the sharpening process mentioned in section 6.1.3 (i.e. removing the host elements Ag and Au) is applied.

columns. However, the formula is not sufficient to predict the interstitial solubility for other pairs of neighboring elements very accurately or unite them. We will therefore work with the extended version

$$a_x \cdot \mathbf{\Gamma}_1 + b_x \cdot \mathbf{\Gamma}_2 + c_x \cdot \mathbf{\Gamma}_3 = \Delta \mathbf{H}_x \quad (6.8)$$

for an interstitial x . In matrix notation it becomes

$$(\mathbf{\Gamma}_1, \mathbf{\Gamma}_2, \mathbf{\Gamma}_3) \cdot [a_x, b_x, c_x]^T = \mathbf{\Gamma}_{(i,j,k)} \cdot [a, b, c]^T = \Delta \mathbf{H}_x, \quad (6.9)$$

in which each of the matrix columns $\mathbf{\Gamma}_i$ can contain a product of maximum three bulk parameters, each of them again representing a column of numbers for the individual host materials.

If m interstitial elements are considered at the same time, equation 6.9 must be extended to

$$\mathbf{\Gamma}_{(i,j,k)} \cdot \begin{pmatrix} a_1 & a_2 & \cdots & a_m \\ b_1 & b_2 & \cdots & b_m \\ c_1 & c_2 & \cdots & c_m \end{pmatrix} = (\Delta \mathbf{H}_1, \Delta \mathbf{H}_2, \cdots, \Delta \mathbf{H}_m) \quad (6.10)$$

and solved. The indices $1, 2, \cdots, m$ number the interstitial elements that are considered and the values a_i, b_i, \cdots, c_m result from a fitting procedure. Each column of the bulk matrix $\mathbf{\Gamma}_{(i,j,k)}$ can contain a product of zero, one, two or three bulk parameters of table 6.3. The number of bulk properties which are used and which gives the form of the predicting equation is marked by the index triple (i, j, k) in $\mathbf{\Gamma}_{(i,j,k)}$. Griessen's equation 3.74 for hydrogen, for example, is the special case $\mathbf{\Gamma}_{(3,0,0)}$ with $\mathbf{\Gamma}_1 = \Delta E \cdot W_d \cdot \sum_j R_j$ (the product is an element-wise product for each host material) and the other two columns are unit vectors of the form $\mathbf{\Gamma}_2 = \mathbf{\Gamma}_3 = \mathbf{1} := [1, 1, 1, \dots, 1]^T$. Therefore, we introduce the general index notation as

$$\begin{aligned} \mathbf{\Gamma}_{(0,0,0)} &= (\mathbf{1}, \quad \mathbf{1}, \quad \mathbf{1}), \\ \mathbf{\Gamma}_{(0,0,1)} &= (\mathbf{1}, \quad \mathbf{1}, \quad \mathbf{t}_1), \\ &\vdots \\ \mathbf{\Gamma}_{(2,3,3)} &= (\mathbf{t}_1 \cdot \mathbf{t}_2, \quad \mathbf{t}_3 \cdot \mathbf{t}_4 \cdot \mathbf{t}_5, \quad \mathbf{t}_6 \cdot \mathbf{t}_7 \cdot \mathbf{t}_8), \\ \mathbf{\Gamma}_{(3,3,3)} &= (\mathbf{t}_1 \cdot \mathbf{t}_2 \cdot \mathbf{t}_3, \quad \mathbf{t}_4 \cdot \mathbf{t}_5 \cdot \mathbf{t}_6, \quad \mathbf{t}_7 \cdot \mathbf{t}_8 \cdot \mathbf{t}_9), \end{aligned}$$

with $\mathbf{1}$ is the unit vector for the column, \mathbf{t}_i is any column of table 6.3 or its inverse, and $\mathbf{t}_i \cdot \mathbf{t}_j$ the element-wise product of table columns i and j . This allows one to consider properties up to their third power, while the unit vector corresponds to a constant in the equation.

The number of possibilities for each column is given with r picks out of $n = 8 \cdot 2 + 1 = 17$ choices (eight descriptors and their inversions), is given by equation

$$\binom{n+r-1}{r} = \frac{(n+r-1)!}{r!(n-1)!}. \quad (6.11)$$

Considering the maximum amount of three picks (with $\mathbf{\Gamma}_{(3,3,3)}$) for each column which are 969

choices, we end up with $969^3 = 909,853,209$ possibilities (equations) which need to be considered and solved for one single interstitial element. Assuming a calculation time of one second for each equation we end up with approx. 10530 days for a single processor run. Reducing the calculation speed to 1/10 seconds, we still need approx three years for the calculation for one single interstitial element for the matrix form $\mathbf{\Gamma}_{(3,3,3)}$. A solution is parallel computing or/and the reduction of the choices n . Both have been performed. The reduction of the choices can be done by using the FA on the physical bulk properties in order to determine correlated parameters. This will be presented below.

6.2.4 FA for physical bulk properties

In order to figure out correlations among the bulk parameters (columns of table 6.3) the Pearson correlation coefficient matrix is calculated and presented (see table 6.4). It can be seen that the

	$W_d^{1/2}$	R_j	ΔE_{F-s}	B_0	f_d	ϵ_d	N_{tot}	N_{occ}
$W_d^{1/2}$	1.00	-0.6	0.41	0.08	-0.07	-0.02	-0.42	-0.11
R_j		1.00	0.25	0.44	0.56	-0.46	0.7	0.57
ΔE_{F-s}			1.00	0.71	0.7	-0.77	0.44	0.66
B_0				1.00	0.27	-0.29	0.18	0.22
f_d					1.00	-0.97	0.85	1.00
ϵ_d						1.00	-0.83	-0.96
N_{tot}							1.00	0.88
N_{occ}								1.00

Table 6.4: Pearson correlation coefficient matrix for the bulk properties. The shading is increasing with the increase of the absolute values. A correlation between N_{occ} , $-\epsilon_d$ and f_d is observed.

fractional d-band filling f_d strongly correlates with the negative average energy of the d-band $-\epsilon_d$ and the number of occupied states N_{occ} (i.e. $f_d \sim -\epsilon_d \sim N_{occ}$). All other studied bulk parameters are much less dependent on each other. In order to determine the best fitting formula and bulk parameters for the interstitial enthalpy prediction it is sufficient to use this value (f_d) or one of the other two correlated parameters instead of using all three of them. This will reduce the dimension of the combinatoric search problem. Instead of having the mentioned 17 choices, n will reduce to 13 (the bulk properties and their inverted values plus unity vector). Using equation 6.11 the possibilities reduce to 455^3 or 94,196,375 choices. The reduction is with one order of magnitude significant. Parallel calculation of the choices is reducing the combinatoric search to less than one hour per interstitial element.

6.2.5 Results for the combinatoric search

In the following, the main results of the combinatoric search are presented and discussed. For this purpose for each combination of descriptors $\mathbf{\Gamma}_{(i,j,k)}$ a fit to the tabulated solutions enthalpies is performed and the Pearson correlation coefficient of the resulting expression and these solution enthalpies is determined.

The physical bulk properties are normalized with respect to the average values (for example the column of the bulk modulus of table 6.3 changes to $\widetilde{B}_0 = B_0/\overline{B}_0$) in order to cancel out the units in the search for the best predictive formula and to compare easily the coefficients in the equations given in the tables C.4-C.12 with each other. These coefficients are interpreted as weight factors, which determine the impact of certain physical bulk parameter(s) on the solution enthalpy. In order to get the right coefficients for the predictive equation we need to multiply these weight factors at the end again with the averaged value(s) of the specific bulk property (or properties).

Starting with hydrogen we obtain the following most important optimizations:

Formula	ΔH function	correlation
$\Gamma_{(0,0,3)}$	$-1.062 + 1.077 \cdot \widetilde{N}_{tot}^{-1} \cdot \widetilde{R}_j \cdot \widetilde{\Delta E_{F-s}}$	0.906
$\Gamma_{(0,1,1)}$	$-1.519 + 1.124 \cdot \widetilde{\Delta E_{F-s}} + 0.424 \cdot \widetilde{R}_j$	0.910
$\Gamma_{(3,3,3)}$	$\left(-0.943 \cdot \widetilde{R}_j^{-1} \cdot \widetilde{B}_0 + 3.352 \cdot \widetilde{f}_d^{-1} \cdot \widetilde{\Delta E_{F-s}} - 2.096 \cdot \widetilde{f}_d^{-1} \cdot \widetilde{W}_d^{1/2} \right) \widetilde{\Delta E_{F-s}}$	0.964

Table 6.5: Selected results of the combinatorial search for the best descriptors for hydrogen (full data given in table C.4). The performance of the predictive formulas is visualized in figure 6.7.

We notice that the best choice for a predictive formula of the structure $\mathbf{\Gamma}_{(0,0,3)}$ is close to the one, which was already suggested in section 3.5 and used by Griessen [114] in 1988. It has the two bulk parameters product $R_j \cdot \Delta E_{F-s}$, i.e. the distance to the neighboring host atoms of the interstitial and the energy difference, in common with this earlier work. This product alone already yields a PCC of 0.884. The third parameter seems to be less important for the predictive power. In the version given in equation 3.74, the parameter $W_d^{1/2}$ is additionally chosen and improves the correlation to 0.905. With this selection, it is ensured that the equation $a_H \Gamma_y + b_H = \Delta H_{H,y}$ has the advantage that the volume dependence is qualitatively correct. Without this additional constraint, the choice of N_{tot}^{-1} as a third parameter has a minor advantage for the predictability of the solution enthalpy.

A slightly higher correlation (0.910 as compared to 0.906) is achieved already with the first two parameters R_j and ΔE_{F-s} , if one uses the structure $\mathbf{\Gamma}_{(0,1,1)}$ of the predictive formula instead, corresponding to the equation $a_H \mathbf{\Gamma}_1 + b_H \mathbf{\Gamma}_2 + c_H = \Delta H_H$. Even the form $\mathbf{\Gamma}_{(1,1,1)}$ does not improve the result. Here, one should notice that ΔE_{F-s} has the higher weight factor a_H . The dominance of this physical property is also reflected in the form $\mathbf{\Gamma}_{(3,3,3)}$. This form yields of course the highest correlation, since it has the highest number of degrees of freedom in order to fit the formula to the targeted solution enthalpy of hydrogen. Comparing the performance of all three results in figure 6.7, one realizes that the form $\mathbf{\Gamma}_{(3,3,3)}$ mainly improves the description of host materials with high solution enthalpies (such as W, Os, Ir). In these cases the band occupation f_d^{-1} seems to play a more dominant role.

The combinatoric search for the carbon solubility predicting equations (tables 6.6 and C.4) yields for the form $\mathbf{\Gamma}_{(0,0,1)}$, $\mathbf{\Gamma}_{(0,0,2)}$, and $\mathbf{\Gamma}_{(0,0,3)}$ the same set of physical parameters as for hydrogen, despite the small modification that \widetilde{N}_{tot}^{-1} instead of \widetilde{N}_{tot} enters $\mathbf{\Gamma}_{(0,0,2)}$. The search further reveals with 0.946 a very high correlation coefficient for the form $\mathbf{\Gamma}_{(0,1,2)}$ where the product of the bulk parameters

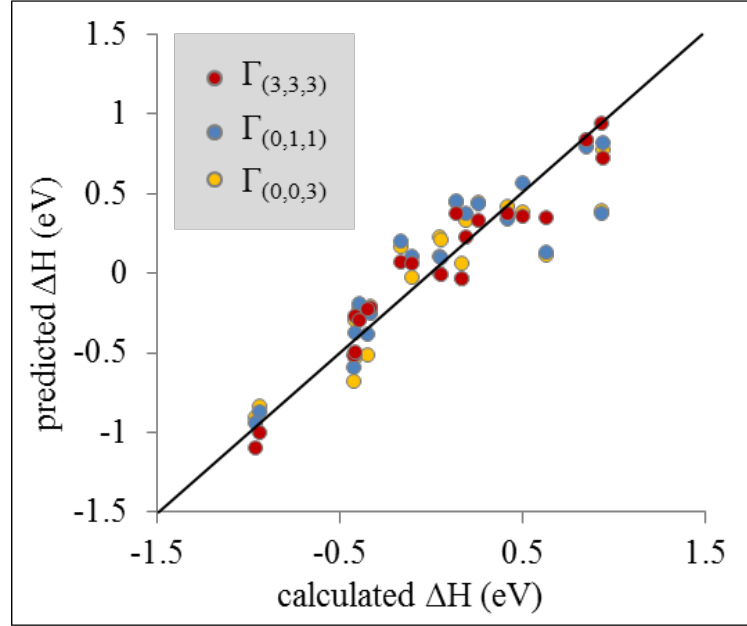


Figure 6.7: Comparison of the solution enthalpy of hydrogen as predicted by the formulas given in table 6.5 and direct calculations. A high correlation (above 0.9) is achieved with two bulk parameters and three fitting values (blue points). Orange points correspond to a predicting equation with exactly the same three bulk parameters as suggested earlier by Griessen [114]. The red points correspond to three fitting values and allows up to nine bulk parameters.

ΔE_{F-s} and R_j remains to be the dominant term. The structure of the formula hardly changes even if one substantially increases the number of allowed physical parameters up to $\Gamma_{(0,3,3)}$. Even for $\Gamma_{(2,3,3)} = \Gamma_{(3,3,3)}$ the terms and the PCC value are only slightly extended. The physical parameter

Formula	ΔH function	correlation
$\Gamma_{(0,0,2)}$	$-2.276 + 2.517 \cdot \widetilde{R}_j \cdot \widetilde{\Delta E_{F-s}}$	0.881
$\Gamma_{(0,1,2)}$	$-4.442 + 0.941 \cdot \widetilde{B}_0^{-1} + 3.694 \cdot \widetilde{R}_j \cdot \widetilde{\Delta E_{F-s}}$	0.946
$\Gamma_{(2,3,3)}$	$-3.956 \cdot \widetilde{W}_d^{1/2} \cdot \widetilde{N}_{tot} + 3.516 \cdot \widetilde{W}_d^{1/2} \cdot \widetilde{R}_j \cdot \widetilde{\Delta E_{F-s}} + 0.752 \cdot (\widetilde{B}_0^{-1})^2 \cdot \widetilde{f}_d$	0.956

Table 6.6: Selected results of the combinatorial search for carbon (full data given in table C.4).

ΔE_{F-s} is also the dominant term for the correlation coefficient for the nitrogen and the oxygen solubility, this time however combined with the band width $\widetilde{W}_d^{-1/2}$ (see tables 6.7 and 6.8). Already in the case of $\Gamma_{(0,0,2)}$ this yields remarkably high PCC values of 0.946 and 0.972, respectively. Even if a larger number of physical parameters is taken into account, these two interstitial elements yield very similar terms in the predictive formulas.

On the other hand, the analysis of the fitted equations for the boron solubility (table C.6)

Formula	ΔH function	correlation
$\Gamma_{(0,0,2)}$	$-6.872 + 6.659 \cdot \widetilde{W_d^{-1/2}} \cdot \widetilde{\Delta E_{F-s}}$	0.946
$\Gamma_{(0,2,2)}$	$-13.542 + 10.58 \cdot \widetilde{W_d^{-1/2}} \cdot \widetilde{\Delta E_{F-s}} + 2.87 \cdot \widetilde{\Delta E_{F-s}^{-1}} \cdot \widetilde{W_d^{1/2}}$	0.970
$\Gamma_{(2,3,3)}$	$-4.453 \cdot \widetilde{f_d^{-1}} \cdot \widetilde{W_d^{1/2}} + 3.219 \cdot \widetilde{N_{tot}^{-1}} \cdot \widetilde{\Delta E_{F-s}} + 0.318 \cdot \widetilde{B_0^{-1}} \cdot (\widetilde{f_d^{-1}})^2$	0.981

Table 6.7: Selected results of the combinatorial search for nitrogen (full data given in table C.5).

Formula	ΔH function	correlation
$\Gamma_{(0,0,2)}$	$-10.476 + 8.576 \cdot \widetilde{W_d^{-1/2}} \cdot \widetilde{\Delta E_{F-s}}$	0.972
$\Gamma_{(0,1,2)}$	$-11.62 + 0.412 \cdot \widetilde{B_0^{-1}} + 9.323 \cdot \widetilde{W_d^{-1/2}} \cdot \widetilde{\Delta E_{F-s}}$	0.974
$\Gamma_{(3,3,3)}$	$\left(0.511 \cdot \widetilde{B_0^{-1}} \cdot \widetilde{f_d^{-1}} + 8.58 \cdot \widetilde{\Delta E_{F-s}} \cdot \widetilde{N_{tot}} - 10.842 \cdot \widetilde{W_d^{1/2}} \cdot \widetilde{N_{tot}} \right) \widetilde{f_d^{-1}}$	0.992

Table 6.8: Selected results of the combinatorial search for oxygen (full data given in table C.7).

reveals that the considered bulk properties of the host materials describe the solubility of boron much less accurately. Already $\Gamma_{(0,0,3)}$, though it has a similar structure as for the other interstitial elements, has a correlation coefficient (0.538) that is much too small for predictions. The observation in section 6.2 that boron requires three factors in a factor analysis is clearly confirmed by these investigations. For example, a comparison of $\Gamma_{(0,1,1)}$ and $\Gamma_{(1,1,1)}$ (i.e. the comparison of a two factors equation vs a three factors equation) shows a substantial increase in the PCC of B (see table 6.9), while it did not improve the description of H and N at all, and only slightly that of C. Also the changes between $\Gamma_{(0,3,3)}$ and $\Gamma_{(3,3,3)}$ are larger for B. Using the highest number of control parameters in $\Gamma_{(3,3,3)}$ yields a reasonable PCC of 0.899, which is, however, still significantly smaller than for the other interstitial elements. This might indicate that the large variety of electronic properties considered in our formalism are less suitable for boron in the dilute limit.

Formula	ΔH function	correlation
$\Gamma_{(0,0,3)}$	$-1.684 + 2.144 \cdot \widetilde{f_d^{-1}} \cdot \widetilde{R_j} \cdot \widetilde{\Delta E_{F-s}}$	0.538
$\Gamma_{(0,1,1)}$	$0.445 + 1.540 \cdot \widetilde{R_j} - 2.067 \cdot \widetilde{W_d^{-1/2}}$	0.498
$\Gamma_{(1,1,1)}$	$0.765 \cdot \widetilde{f_d^{-1}} + 3.406 \cdot \widetilde{R_j} - 4.234 \cdot \widetilde{W_d^{-1/2}}$	0.644
$\Gamma_{(3,3,3)}$	$\left(-3.207 \cdot \widetilde{N_{tot}} + 5.861 \cdot \widetilde{R_j} \cdot \widetilde{\Delta E_{F-s}} - 2.915 \cdot \widetilde{B_0} \cdot \widetilde{f_d} \right) \widetilde{B_0}$	0.899

Table 6.9: Result of the combinatorial search for boron (full data given in table C.6).

For the interstitial element fluorine already the form $\mathbf{\Gamma}_{(0,0,1)}$ reveals with 0.904 a high correlation coefficient. The form $\mathbf{\Gamma}_{(0,0,2)}$ with the element-wise product of the d-band center f_d and the bulk modulus B_0 , shows with 0.964 a very high correlation between the predicting formula and the calculated enthalpies (see table 6.10). The bulk modulus remains the key quantity for fluorine also for the other forms of $\mathbf{\Gamma}$.

Formula	ΔH function	correlation
$\Gamma_{(0,0,2)}$	$-4.412 + 4.038 \cdot \widetilde{B}_0 \cdot \widetilde{f}_d$	0.964
$\Gamma_{(0,1,1)}$	$-7.038 + 2.933 \cdot \widetilde{f}_d + 4.135 \cdot \widetilde{B}_0$	0.978
$\Gamma_{(1,2,3)}$	$-9.162 \cdot \widetilde{N}_{tot} + 7.834 \cdot \widetilde{W}_d^{-1/2} \cdot \widetilde{\Delta E_{F-s}} + 1.286 \cdot (\widetilde{R}_j)^2 \cdot \widetilde{B}_0$	0.987

Table 6.10: Selected results of the combinatorial search for fluorine (full data given in table C.8).

It is again the bulk modulus which is the best parameter to predict the solution enthalpy of helium and neon as well. The predicting form of $\mathbf{\Gamma}_{0,0,1}$ is sufficient to achieve a correlation of 0.937 for helium and 0.951 for neon (see tables 6.11 and 6.12).

Formula	ΔH function	correlation
$\Gamma_{(0,0,1)}$	$1.354 + 2.809 \cdot \widetilde{B}_0$	0.937
$\Gamma_{(0,1,1)}$	$1.022 + 2.512 \cdot \widetilde{B}_0 + 0.628 \cdot \widetilde{f}_d$	0.956
$\Gamma_{(2,3,3)}$	$5.593 \cdot \widetilde{W}_d^{-1/2} \cdot \widetilde{\Delta E_{F-s}} - 3.449 \cdot \widetilde{f}_d \cdot (\widetilde{N}_{tot})^2 + 2.021 \cdot \widetilde{W}_d^{-1/2} \cdot \widetilde{R}_j \cdot \widetilde{f}_d$	0.994

Table 6.11: Selected results of the combinatorial search for helium (full data given in table C.9).

Formula	ΔH function	correlation
$\Gamma_{(0,0,1)}$	$1.964 + 5.452 \cdot \widetilde{B}_0$	0.951
$\Gamma_{(0,1,1)}$	$0.929 + 5.043 \cdot \widetilde{B}_0 + 1.443 \cdot \widetilde{R}_j$	0.964
$\Gamma_{(3,3,3)}$	$11.462 \cdot \widetilde{W}_d^{-1/2} \cdot \widetilde{N}_{tot}^{-1} \cdot \widetilde{f}_d + 3.353 \cdot \widetilde{R}_j \cdot (\widetilde{\Delta E_{F-s}})^2 - 6.711 \cdot \widetilde{N}_{tot}^{-1} \cdot (\widetilde{f}_d)^2$	0.991

Table 6.12: Selected results of the combinatorial search for neon (full data given in table C.10).

6.2.6 Conclusions from the combinatoric search

The discussion of the previous section can be used for a classification of the interstitial elements. Using the form $\mathbf{\Gamma}_{(0,0,2)}$ as an indicator, the energy difference ΔE_{F-s} is the dominant bulk parameter for the four interstitial elements H, C, N, and O, whereas for the three interstitial elements F, He,

and Ne the dominant bulk parameter is the bulk modulus B_0 . Therefore, the combinatoric search confirms an observation already made in section 5.5.3 that chemical effects are most important for the mentioned first four interstitial elements and that elastic effects dominate the solubility of the noble gases and fluorine (figure 5.22). B has an exceptional behavior that cannot be associated with one of the two groups.

Formula	ΔH function	correlation
$\Gamma_{(0,0,2)}$	$-0.982 + 0.971 \cdot \widetilde{R}_j \cdot \widetilde{\Delta E_{F-s}}$	0.916
$\Gamma_{(0,1,2)}$	$-0.779 - 0.088 \cdot \widetilde{B}_0^{-1} + 0.861 \cdot \widetilde{R}_j \cdot \widetilde{\Delta E_{F-s}}$	0.939
$\Gamma_{(2,3,3)}$	$\left(0.074 \cdot \widetilde{W_d^{-1/2}} \cdot \widetilde{B}_0^{-1} - 0.708 \cdot \widetilde{W_d^{1/2}}\right) \widetilde{\Delta E_{F-s}^{-1}} + 0.626 \cdot \widetilde{N_{tot}^{-1}} \cdot \widetilde{R}_j \cdot \widetilde{\Delta E_{F-s}}$	0.948

Table 6.13: Selected results of the combinatorial search for the group I elements containing hydrogen, carbon, nitrogen, and oxygen (full data given in table C.11). The coefficients shown above corresponds to the interstitial element hydrogen, whereas the correlation value is for the average of all four interstitial elements. The unnormalized coefficients also for the remaining three interstitial elements are given in figure 6.8.

If one takes the physical similarity for the solution enthalpy equation of the form $\mathbf{\Gamma}_{(0,0,2)}$ seriously, then the determination of the bulk parameters in the combinatorial research should be performed for all elements in the same group simultaneously. The result of this procedure is shown in tables 6.13 / C.11 for the group consisting of the interstitial elements H, C, N, and O and in tables 6.14 / C.12 for the group consisting of the elements F, He and Ne.

Formula	ΔH function	correlation
$\Gamma_{(0,0,1)}$	$-5.492 + 5.523 \cdot \widetilde{B}_0$	0.922
$\Gamma_{(0,1,1)}$	$-7.038 + 4.135 \cdot \widetilde{B}_0 + 2.933 \cdot \widetilde{f_d}$	0.962
$\Gamma_{(0,1,2)}$	$-6.25 + 1.599 \cdot \widetilde{B}_0 + 4.5 \cdot \widetilde{R}_j \cdot \widetilde{\Delta E_{F-s}}$	0.966
$\Gamma_{(0,3,3)}$	$-6.295 + 6.394 \cdot \widetilde{N_{tot}^{-1}} \cdot \widetilde{R}_j \cdot \widetilde{\Delta E_{F-s}} - 0.196 \cdot \widetilde{B}_0^{-1} \cdot \widetilde{\Delta E_{F-s}} \cdot \widetilde{f_d}$	0.978

Table 6.14: Selected results of the combinatorial search for the group II elements containing fluorine, helium and neon (full data given in table C.12). The coefficients shown above corresponds to the interstitial element fluorine. The unnormalized coefficients are shown in figure 6.9.

The form $\mathbf{\Gamma}_{(0,0,2)}$ for the group I elements contains besides ΔE_{F-s} the physical bulk parameter R_j , which is imposed by the behavior of hydrogen and carbon. Correspondingly, the PCC value is unchanged for these two elements, but it also decreases only insignificantly for nitrogen and oxygen. The average correlation is already for this simple form above 0.9. An interesting observation is the fact that the form $\mathbf{\Gamma}_{(0,0,3)}$, which has also been used by Griessen for hydrogen [114] yields the same structure as the form $\mathbf{\Gamma}_{(0,0,2)}$ (table C.11). There is no additional physical parameter that, when multiplied with $R_j \cdot \Delta E_{F-s}$, will increase the correlation coefficient for the group I of interstitial

elements together. The structure of the predictive formula remains robust, even if some additional terms are added. For example, also the form $\Gamma_{(1,1,2)} = \Gamma_{(0,1,2)}$ contains the element-wise product $R_j \cdot \Delta E_{F-s}$ and as an additional summand the inverse bulk modulus $B_0^{(-1)}$, which is imposed by the predictive equation of the carbon solubility. Even in the form $\Gamma_{(0,1,2)}$ the element-wise product $R_j \cdot \Delta E_{F-s}$ is still the dominating term. The predictive power of the derived formulas of the form $\Gamma_{(0,0,2)}$ and $\Gamma_{(0,1,2)}$ are shown in figure 6.8.

Analyzing table C.12 for the group II elements, the average correlation coefficient is already high for all three elements using only the bulk modulus B_0 . This is due to the fact, that the noble gases reach a high correlation coefficient whereas the correlation coefficient for the element fluorine lies below 0.9. However the correlation coefficient of fluorine increases drastically by including the d-band center as a parameter f_d . This is shown in figure 6.9. The bending of the fluorine data points (blue circles in left figure) disappears when including f_d (right figure). Instead of using the form $\Gamma_{(0,1,1)}$ with the bulk properties B_0 and f_d , one can also use the form $\Gamma_{(0,1,2)}$. This version preserves the bulk property B_0 of the group II interstitial elements and combines it with the element-wise product of ΔE_{F-s} and R_j , which was for dominating term for the group I elements. The such obtained average correlation coefficient of 0.966 is very close to the correlation coefficient of the form $\Gamma_{(0,1,1)}$. An interesting study could be the analysis of the correlation coefficients for mixed bulk properties, such as the correlation of the element-wise product $R_j \cdot \Delta E_{F-s}$ with f_d , which has been skipped here for brevity.

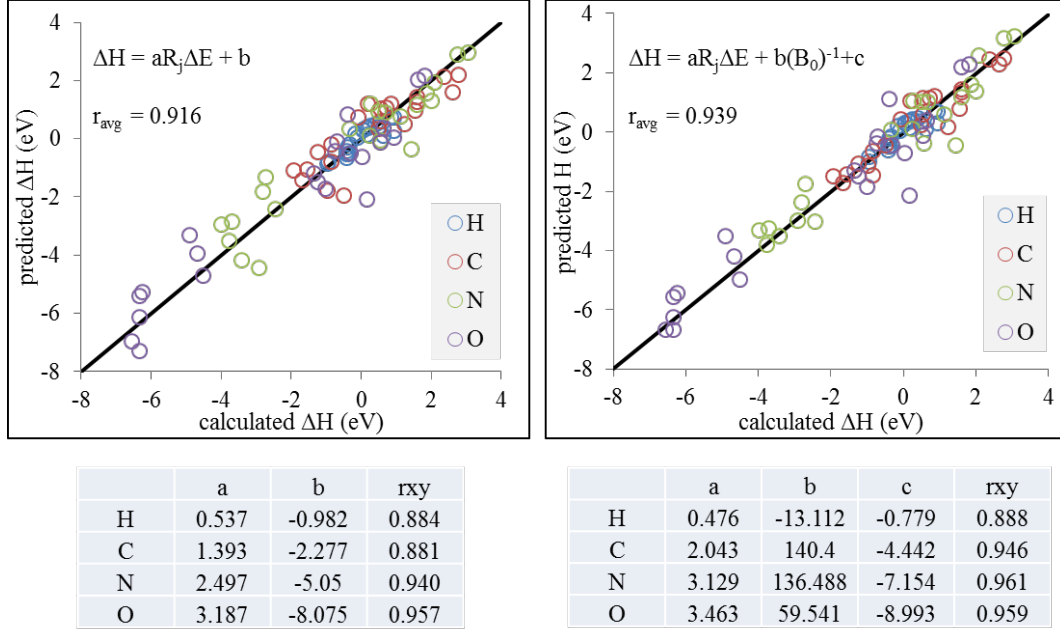


Figure 6.8: The solution enthalpy of H, C, N, and O: Prediction vs calculations. The fitted coefficients here are related to the physical bulk properties which are not normalized as shown in table 6.3. The right figure contains in addition a bulk parameter and a fitting parameter more than the left figure, which improves the carbon correlation coefficient.

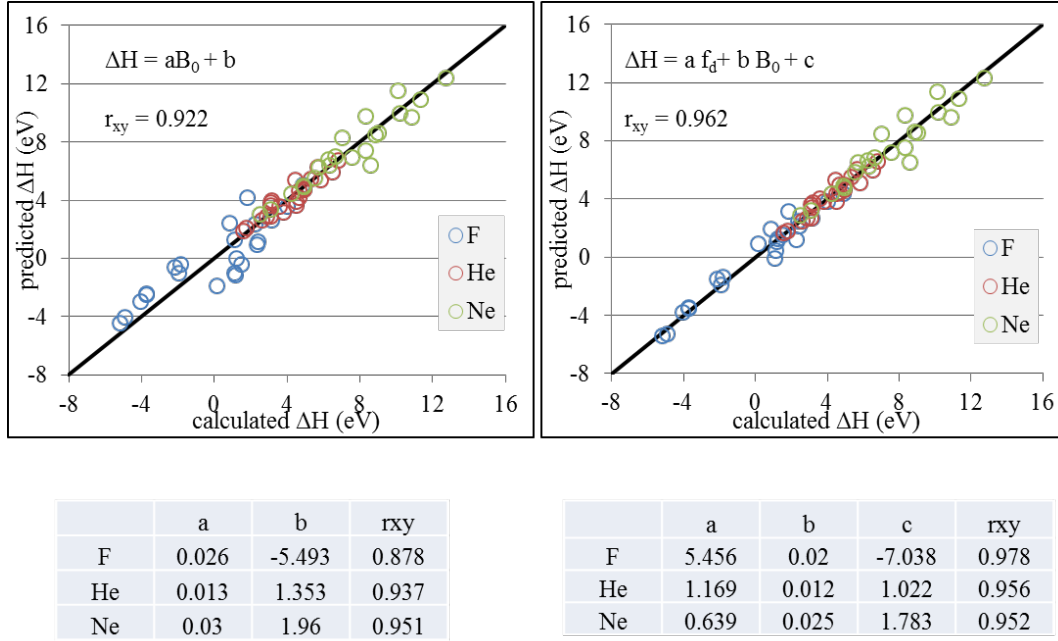


Figure 6.9: The solution enthalpy of F, He and Ne: Prediction vs calculation. The fitted coefficients here are related to the physical bulk properties which are not normalized as shown in table 6.3. The correlation coefficient of fluorine increases drastically by using a second parameter (i.e. f_d) beside the bulk modulus B_0 .

6.2.7 Master curve for interstitial solubility

The discussion above yields to the result that the group of the interstitial elements H, C, N, and O should be distinguished from the other interstitial elements considered in this work. The application of the FA on the solution enthalpies of these four interstitial elements yields the eigenvalues (3.708, 0.188, 0.002, 0.101). The application of the Kaiser-Guttman criterion implies that only one single (common) factor is necessary in order to determine the solution enthalpies of these elements with a minor error. The quality of the reduced component system can be again calculated by taking the difference of the full correlation matrix of the four elements and the calculated correlation matrix of the reduced system, i.e.

$$\mathbf{R}_{\text{red}^1} - \mathbf{R} = \begin{pmatrix} -0.144 & 0.0377 & 0.059 & 0.039 \\ 0.038 & -0.071 & -0.003 & 0.038 \\ 0.059 & -0.003 & -0.028 & -0.025 \\ 0.039 & 0.038 & -0.025 & -0.049 \end{pmatrix}. \quad (6.12)$$

The finding is in perfect agreement with the identification of a single descriptor in the form $\mathbf{\Gamma}_{(0,0,3)}$ containing the product of the two pure bulk properties ΔE_{F-s} and R_j is sufficient to predict the solution enthalpy of these interstitial elements with high precision. This argument can be interpreted as a proof of the concept applied here to predict equations for the solution enthalpies.

The insights obtained in the analysis can be used to construct a master curve for the solubility

of the interstitial elements H, C, N, and O. We use the fitting parameters in the left figure 6.8 to construct contour lines for vanishing solution enthalpy in a diagram where the bulk properties ΔE_{F-s} and $\sum R_j$ form the axes. Comparing it with the originally calculated solubilities in table C.1 it shows an amazing agreement. This leads to the conclusion that the even more simple product $\Delta E_{F-s} \cdot \sum R_j$ than the one found by Griessen for hydrogen is also applicable (with different fitting parameters) for the other interstitial elements C, N, and O.

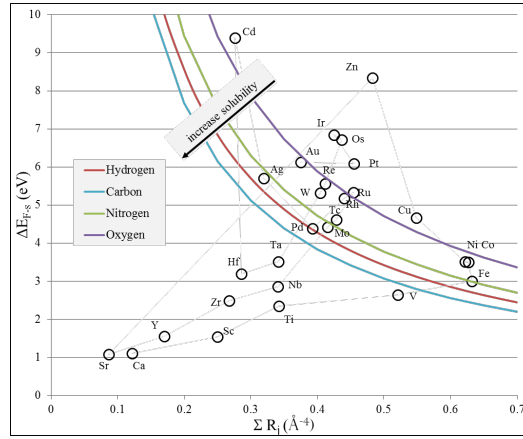


Figure 6.10: Solubility contour lines for the interstitial elements in 3d, 4d and 5d TMs with respect to the descriptors ΔE_{F-s} and $\sum R_j$. The points on the line corresponding to zero solution enthalpy. Left to the line the corresponding interstitial element is predicted to have a negative solution enthalpy and vice-versa.

Chapter 7

Summary and Outlook

A main goal of the present thesis was the discovery of new insights into chemical and physical causes for interstitial solubility in transition metals. We achieved this goal in a three-fold approach that includes (i) the development of tools to perform high-throughput (HT) calculations, (ii) a deep analysis of physical concepts governing hydrogen solubility for a limited set of data, and (iii) the creation of large databases and its analysis with tailored statistical concepts. As a result of this approach we obtained a very compact and previously unknown analytic expression for the prediction of interstitial solubility based on a couple of bulk parameters of the host materials only. In this spirit the current work is a true example of knowledge discovery in databases (KDD).

Setup of the database

The results obtained from the ab initio study of the interstitial element hydrogen in 4th row elements including 3d transition metals (TM) motivated us to extend the study to 5th row (including 4d TMs) and 6th row elements (including 5d TMs). We further learned from the hydrogen investigation that a constraint to non-magnetic fcc structure is helpful and meaningful. Furthermore, the interstitial elements boron, carbon, nitrogen, oxygen, fluorine and the noble gases helium and neon have been taken into account. The host elements are first calculated in their ground state structure and in a second step under comparable conditions in order to rule out structural effects that might influence the interstitial solubility. For this reason all host elements are assumed to be in non-magnetic fcc structure. Considering all these extensions together (including convergence checks) we anticipated the demand of a huge amount of data production, data storage and its analysis. To cope with these challenges of the present high-throughput (HT) study we designed and developed new tools.

Tools to perform HT calculations

Following the concept of the KDD process discussed in chapter 4, we developed a workbench called pyCMW. It is characterized by a variety of features including the independence of platforms, the support of various external programs (VASP, LAMMPS, and KMC) as well as a graphical user interface controlled by the front-end. For the purpose of the present study mainly the concepts implemented in the back-end are decisive. In the first place they allowed us to set up and admin-

istrate the roughly one million individual VASP calculations that entered this HT study. For this purpose, we had to take care of reliable communication with the hardware infrastructure (compute clusters), intelligent storage concepts (using the HDF5 file format) and easy tools for combining and post-processing data.

The last point is already required for standard tasks like Murnaghan fits and the construction of dynamical matrices, which have been heavily used in this thesis. More important was, however, the statistical evaluation like the determination of Pearson correlation coefficients (PCC) and factor analysis. To meet these challenges, the usage of python as a programming language was very helpful, since it allowed the straight-forward implementation of already existing libraries. Furthermore, the concept of grouping calculations into projects, as implemented in pyCMW, greatly simplified the handling of data. Finally the flexibility of the workbench due to its object-oriented and modular structure was decisive, since the way how different sets of data should be combined to determine the best expression for the solution enthalpy, was not clear at the time of programming.

Finally, the graphical tools implemented in the front-end form an important part of the analysis. Though numbers like the PCC eventually quantify the degree of correlation between different physical properties, the knowledge discovery, i.e. the physical understanding of relations and the detection of special properties of individual elements (such as the interstitial boron) cannot be automatized. It can only be achieved by the interaction with a scientist. For the latter, the importance of diagrams cannot be overstated.

Physical concepts governing hydrogen solubility

We first identified chemical trends for the solution enthalpy ΔH of the interstitial element hydrogen in the 4th row elements. The study was restricted to (i) 3d transition metals and the neighboring elements K, Ca, Ge, and Ga, (ii) the fcc crystal structure and (iii) non-magnetic configurations. Using these constraints, we were able to identify several rules for the hydrogen solubility in transition metals:

1. A universal dependence of the hydrogen solution enthalpy on the lattice constant of the host metal was revealed. The universality becomes most apparent, if tensely or compressively strained materials are considered.
2. The solution enthalpy results from a competition between getting sufficient space to incorporate hydrogen (i.e. a preference for large lattice constants) and the formation of strong hydrogen-metal bonds (favoring smaller atomic distances). In the present thesis, we identified an optimal spherical radius for interstitial H of approximately 0.7 Å in the octahedral site (0.4 Å in the tetrahedral site, respectively).
3. An fcc-lattice constant of 4.6 Å turned out to be characteristic for several atomistic mechanisms. In particular at this lattice constant, no relaxation of matrix atoms is needed upon hydrogen incorporation at the octahedral site.
4. These observations motivate a decomposition of the hydrogen solution enthalpy in a strain induced part ΔE_{strain} (related to the lattice relaxations) and a remaining chemical part ΔE_{chem} .

The such defined chemical part, ΔE_{chem} , contains the energy related to the formation of chemical bonds. It covers the largest part of the hydrogen solution enthalpy. Accordingly, its volume dependence also shows the universal behavior observed for the whole solution enthalpy.

5. Investigating charge density differences at the hydrogen position revealed two classes of elements: In one class only a small part of the electronic charge of the matrix material is locally bound at the hydrogen position. This is a consequence of the limited ability of these materials to screen the proton. The resulting reduction in ionic character explains the significantly higher endothermic hydrogen solution enthalpy.

The insights of the study of hydrogen in 4th row elements motivated us to extend the research to 5th row and 6th row host elements and to consider additional interstitial elements including the elements boron, carbon, nitrogen, oxygen, fluorine, helium, and neon. The previously revealed trends for hydrogen were replicated for the interstitial elements boron, carbon, nitrogen, and oxygen, though they were less clearly visible in these cases. The trends did not show up for the noble gases and fluorine, which have their ΔH maximum at half-filled d-shells (similar to the behavior of the bulk modulus w.r.t. d-band filling).

Knowledge discovery from our databases

The PCC analysis of the solution enthalpy versus bulk modulus and Griessen's ΔE term revealed that the elements F, He, Ne are strongly correlating with the bulk modulus, which is an evidence that the solubility for these elements is mainly determined by elastic contributions. With increasing the bulk modulus the solubility is decreasing and in the case of Ne two times higher than the element He. The solution enthalpy of hydrogen is mainly influenced by an chemical effect, which is in excellent agreement with the study of H in 4th row elements before.

The PCC analysis of the solution enthalpies among the interstitial elements revealed high correlation coefficients between the interstitial solubilities of neighboring interstitial elements in the PES (having a single electron difference only), such as C and N, N and O, or F and Ne or elements in the same column of the PES (having same number of electrons in the outer shells), such as He and Ne. A surprisingly high correlation coefficient between the elements H and N, H and C, and H and O was observed. N is approximately 4 times more soluble than H, and C approximately 2.5 times more soluble than H.

The Factor Analysis (FA) of the interstitial element solubilities revealed that two latent factors are necessary to describe the solution enthalpy of almost all interstitial elements except boron. To describe boron interstitial solubility a third latent factor appeared. In order to find a predictive model and parameters for the interstitial solubilities a combinatorial search has been applied. It turned out, that fitting is best performed, if the interstitial elements are divided into two groups. The first group consists of the interstitial elements H, C, N, and O, and the second group consists of F, He, and Ne. It turned out that for the first group the formula

$$\Delta H = a\Delta E \cdot \sum_j R_j + b \quad (7.1)$$

is sufficiently accurate to draw a treasure map for the interstitial solubilities. This is shown in

figure 6.10. This formula reflects our central finding that predictive model should reflect both the chemical nature, as it is done by ΔE , and the mechanical nature, as done by $\sum_j R_j$. Beyond this, a third parameter, as used in earlier work by Griessen [114], is not required, unless another target property like the formation volume should also be captured.

Outlook

We believe that the achieved general understanding of fundamental rules for the incorporation of interstitial elements and the obtained knowledge of chemical trends are very important for the design of materials in which dilute interstitial concentration have severe effects on the functionality and/or structural performance.

Practical applications require, however, an extension of the investigations in several directions. First, the current approach is not free of approximations. In particular, the choice of the exchange correlation-functional within the DFT calculations can have a strong impact on the final results for the solubility. We believe that the chemical trends will not change, if another functional is chosen. Nevertheless, an estimate of the predictive power could be achieved by comparing various functionals. The obtained insights might also be helpful, to develop tight-binding parametrizations, which allow the investigation of more complex structures.

Second, the host materials in reality are typically not pure elements, but multicomponent alloys. It is an assumption of the present study that the impact, for example, of Cr in Ni on the H solubility is qualitatively already covered, if the solubility in pure Cr is compared to that in pure Ni, provided that both elements are treated in the same crystal structure as the alloy. The extension to multicomponent systems needs, however, more systematic investigations, starting from alloying elements in the dilute limit and proceeding to concentrated ordered and disordered alloys.

Third, the solution of elements, in particular if incorporated as an interstitial, may not be well described by considering perfect single crystals. In applications, materials have a complicated microstructure, consisting of several grains, point defects and extended defects. For all these microstructure features similar high throughput investigation can be performed or are already on the way. Since they are computationally even more demanding, the existence of efficient simulation tools and analysis concepts, as developed in this work, is highly important.

Finally, one rarely wants to optimize a single quantity, like the interstitial solubility. Typically, other parameters, such a electrical conductivity, phase stability, solution hardening, stacking fault energy, etc. should be controlled at the same time. Therefore, the correlation between different target properties should be systematically investigated. This can be achieved by extending the concepts of knowledge discovery that turned out to be so valuable for the physical questions presented in this work.

Appendix A

Correlation of Griessen's predictive equation

The correlation between the rearranged equation 5.7 versus the d-bandwidth W_d is presented below

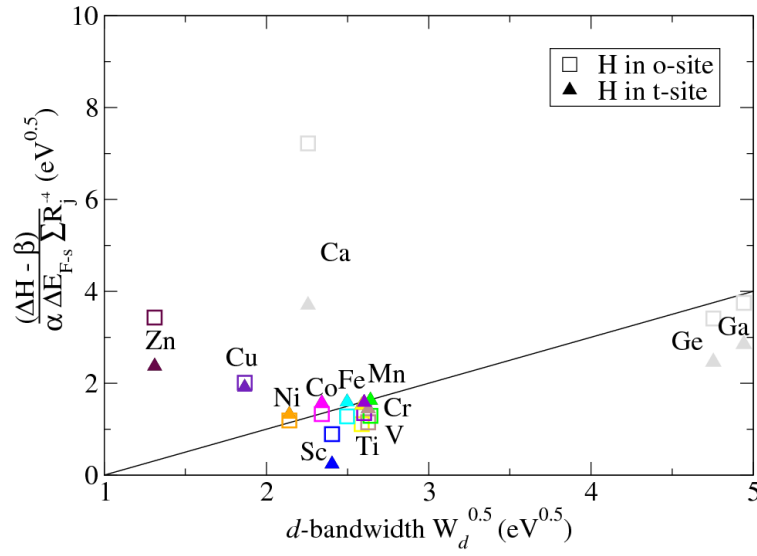


Figure A.1: The correlation between the d-bandwidth W_d and Griessen's rearranged equation 5.7.

Appendix B

Study of interstitial elements

B.1 Hydrogen solution enthalpy vs lattice constant

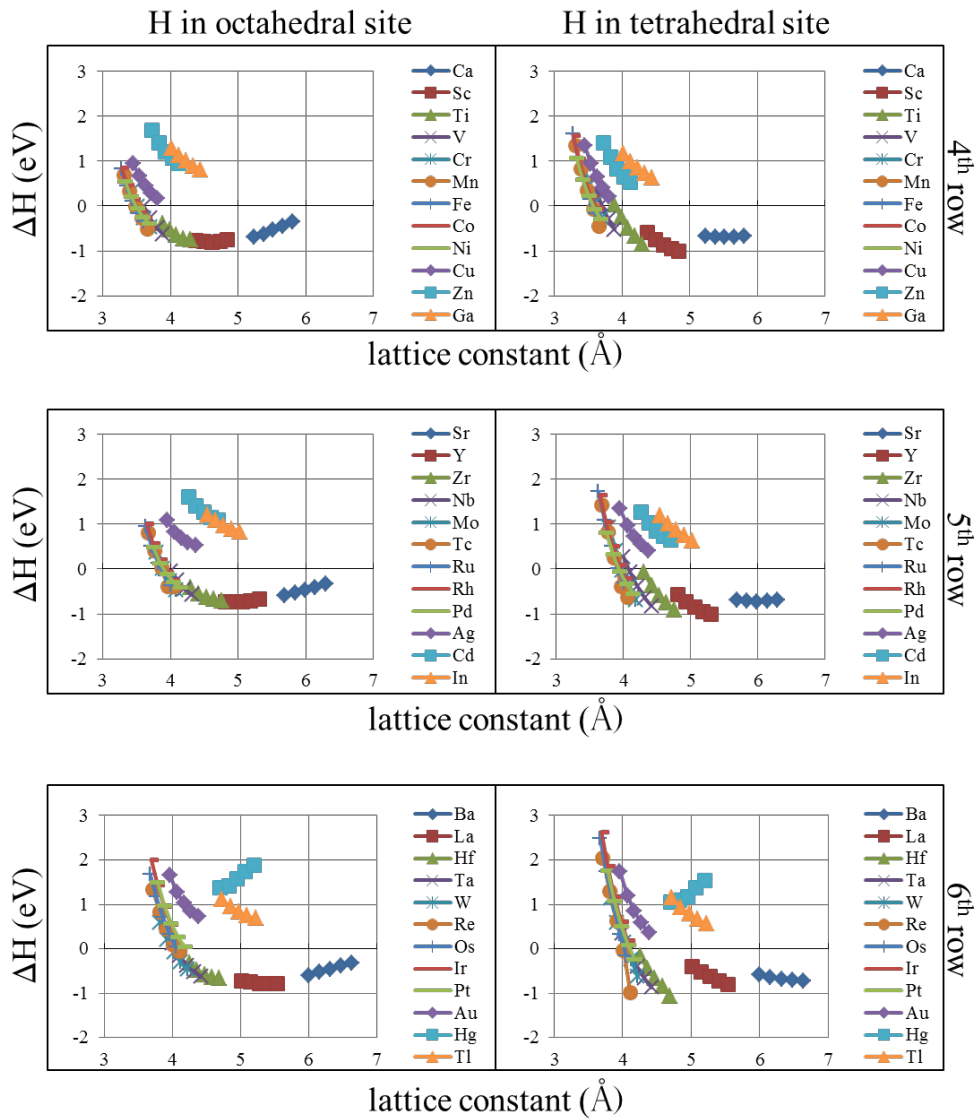


Figure B.1: Solution enthalpy of hydrogen in 4th, 5th, and 6th row elements in the octahedral (left) and in the tetrahedral (right) site.

B.2 Derivative of hydrogen solution enthalpy

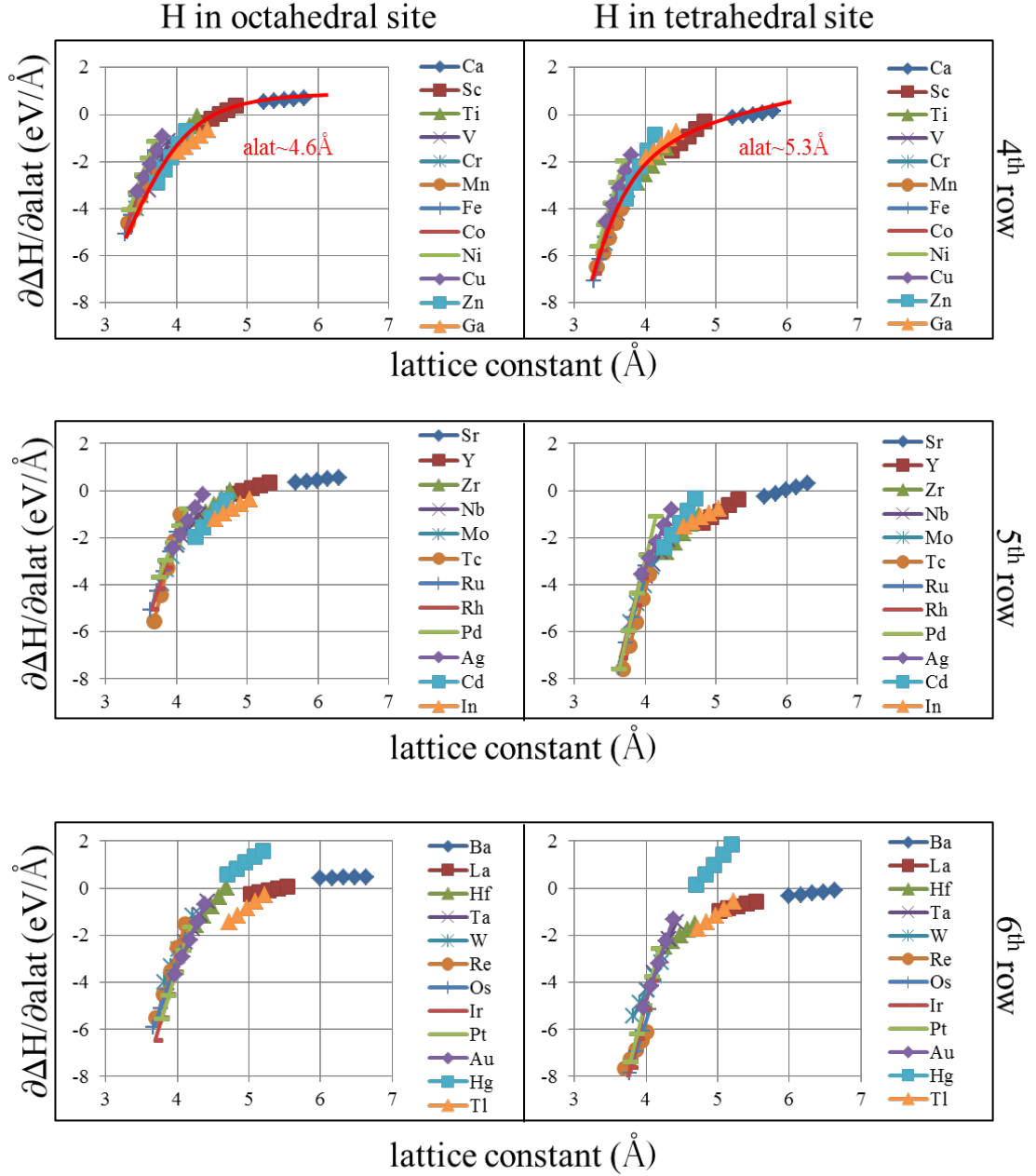


Figure B.2: Derivative of the solution enthalpy of hydrogen in 4th, 5th, and 6th row elements in the octahedral (left) and in the tetrahedral (right) site.

B.3 Interstitial solution enthalpies vs lattice constant

B.3.1 Boron solution enthalpy vs lattice constant

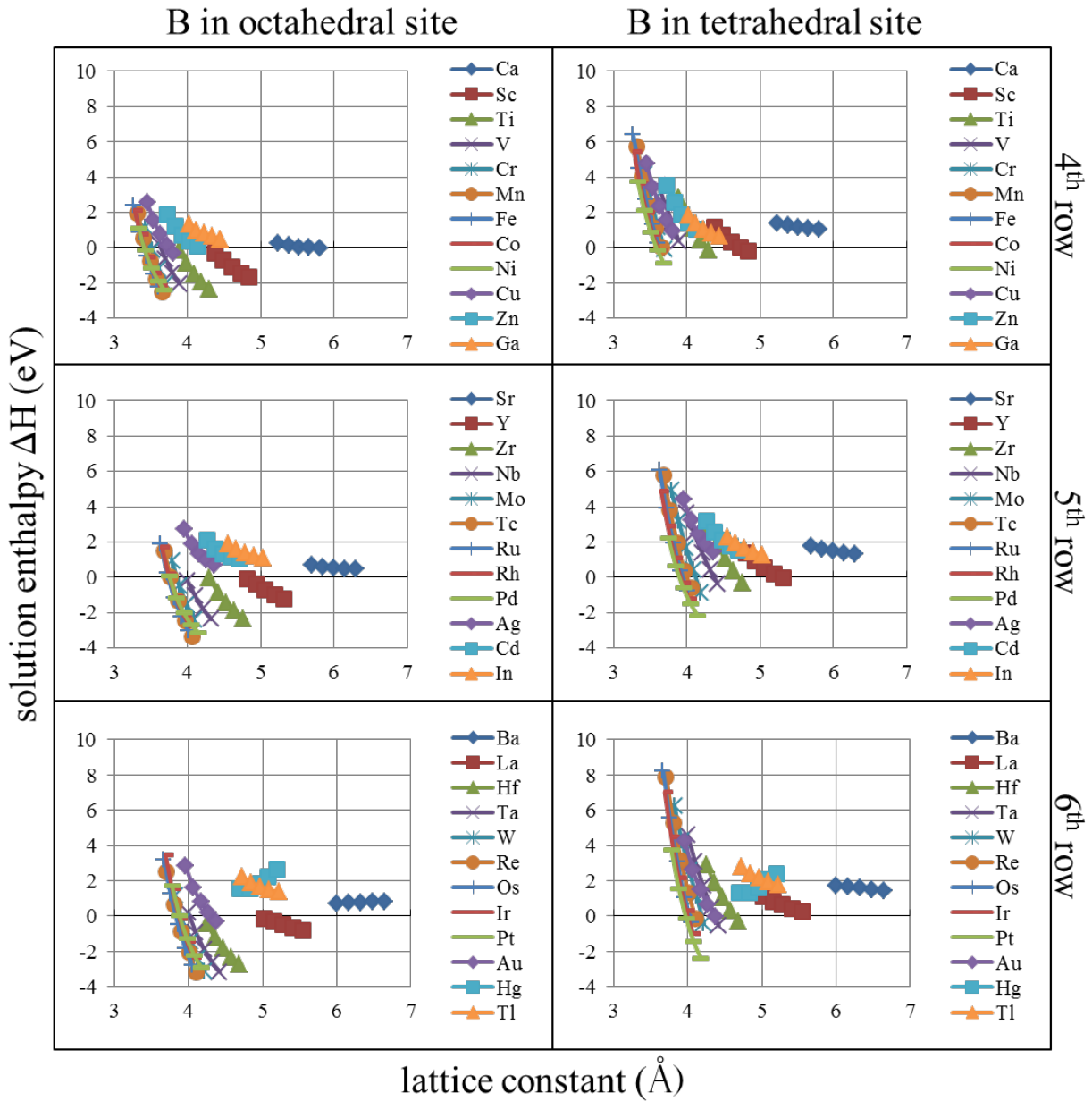


Figure B.3: Solution enthalpy of boron in 4th, 5th, and 6th row elements in the octahedral (left) and in the tetrahedral (right) site.

B.3.2 Carbon solution enthalpy vs lattice constant

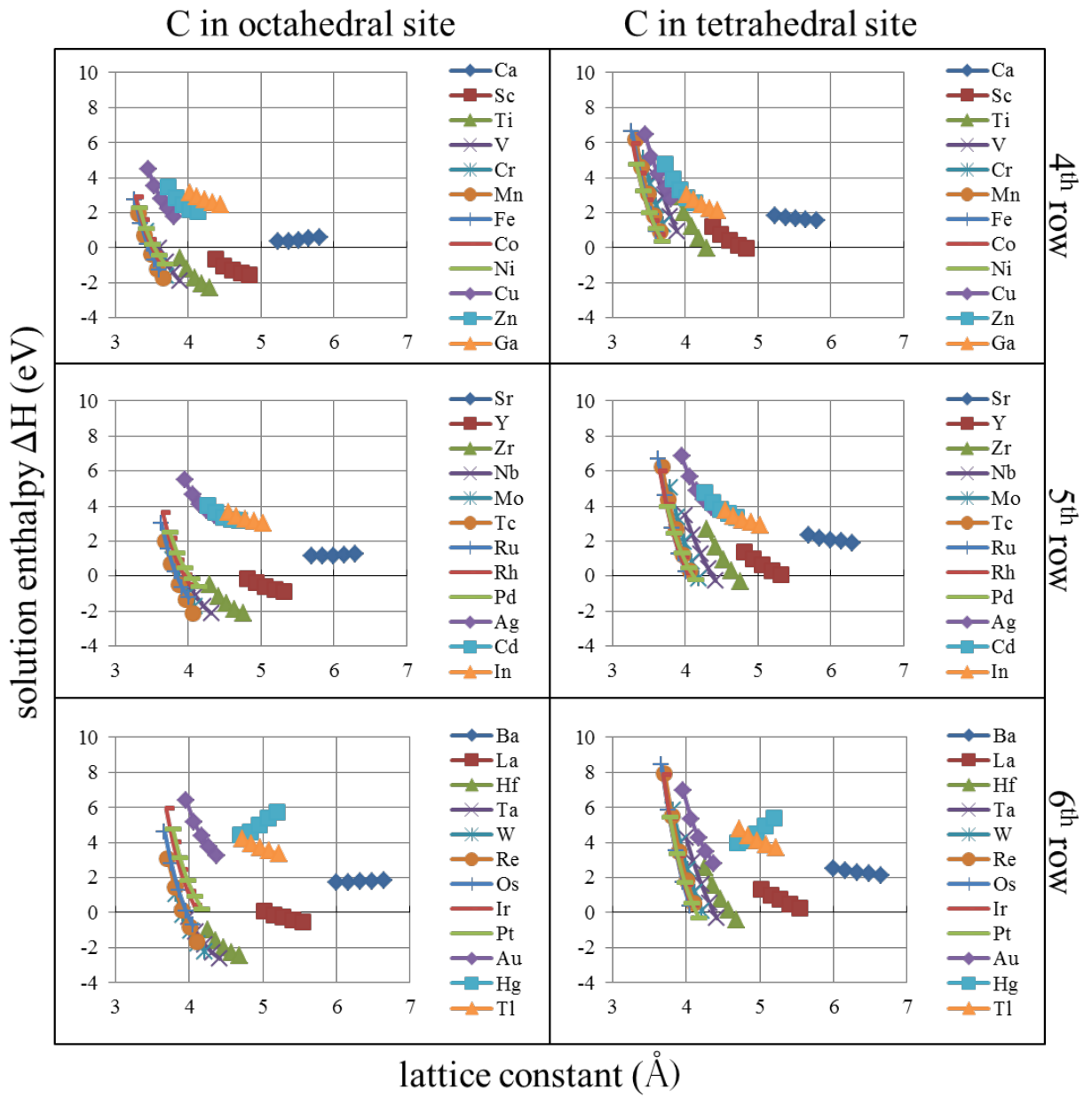


Figure B.4: Solution enthalpy of carbon in 4th, 5th, and 6th row elements in the octahedral (left) and in the tetrahedral (right) site.

B.3.3 Nitrogen solution enthalpy vs lattice constant

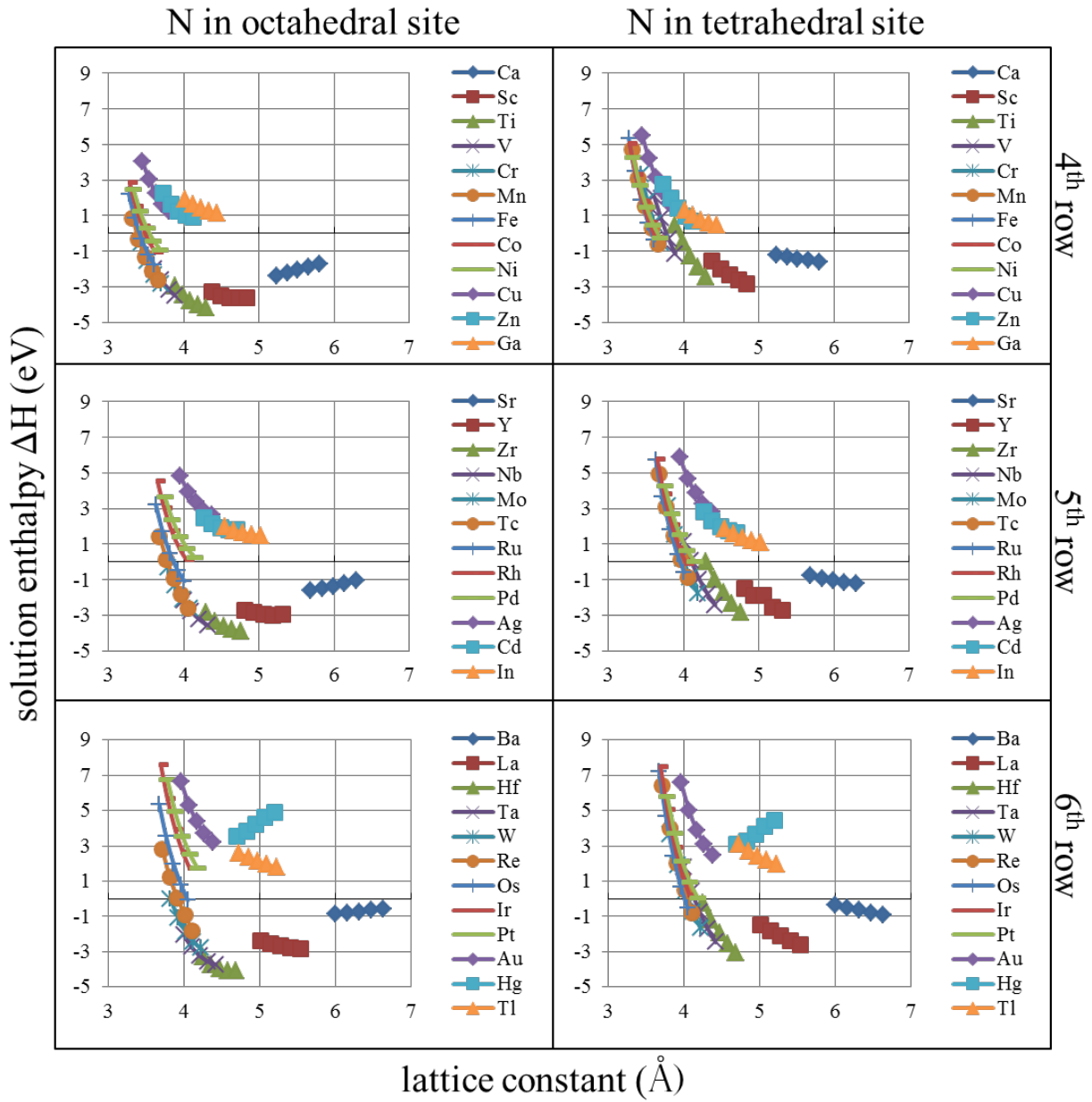


Figure B.5: Solution enthalpy of nitrogen in 4th, 5th, and 6th row elements in the octahedral (left) and in the tetrahedral (right) site.

B.3.4 Oxygen solution enthalpy vs lattice constant

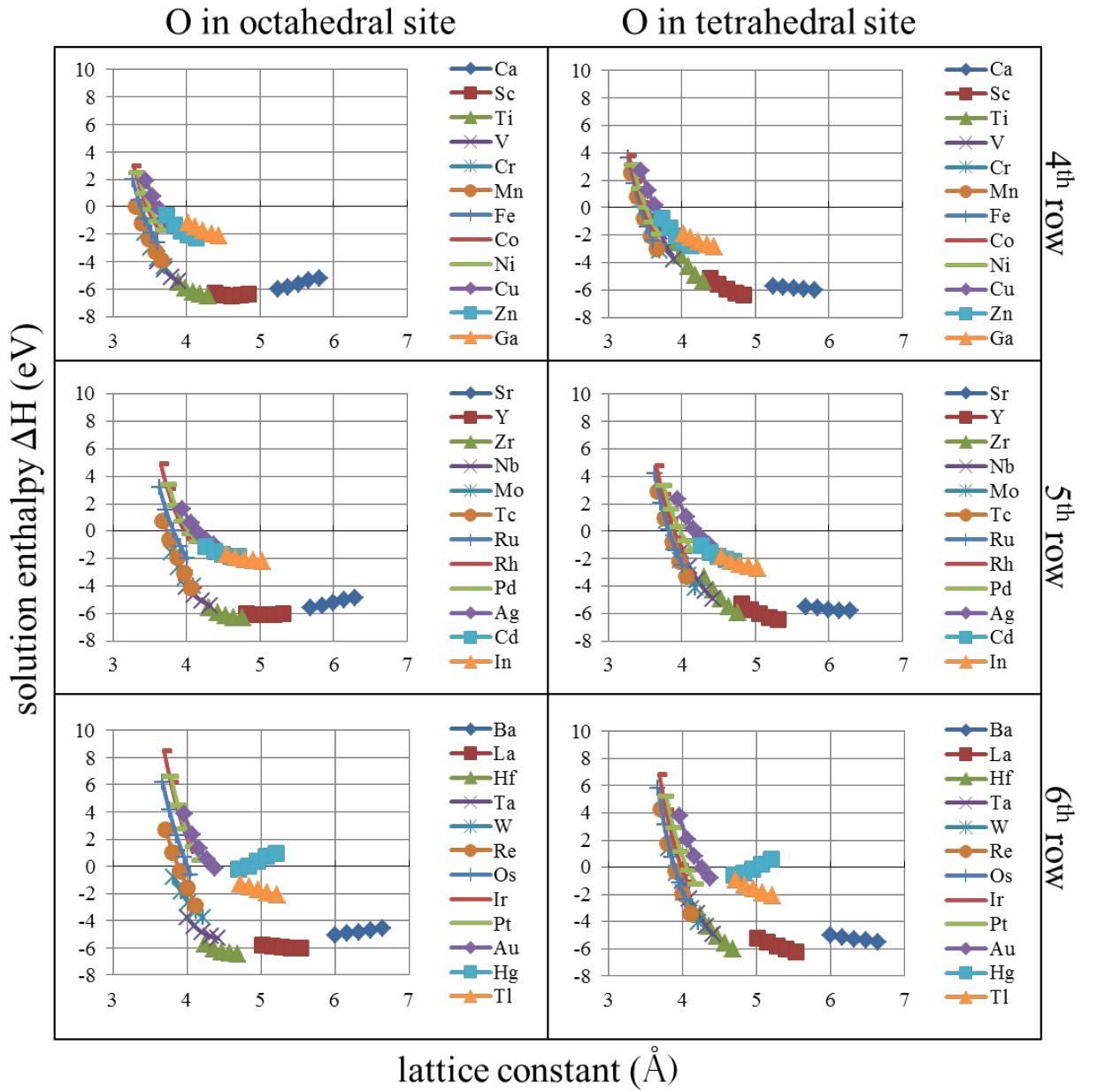


Figure B.6: Solution enthalpy of oxygen in 4th, 5th, and 6th row elements in the octahedral (left) and in the tetrahedral (right) site.

B.3.5 Fluorine solution enthalpy vs lattice constant

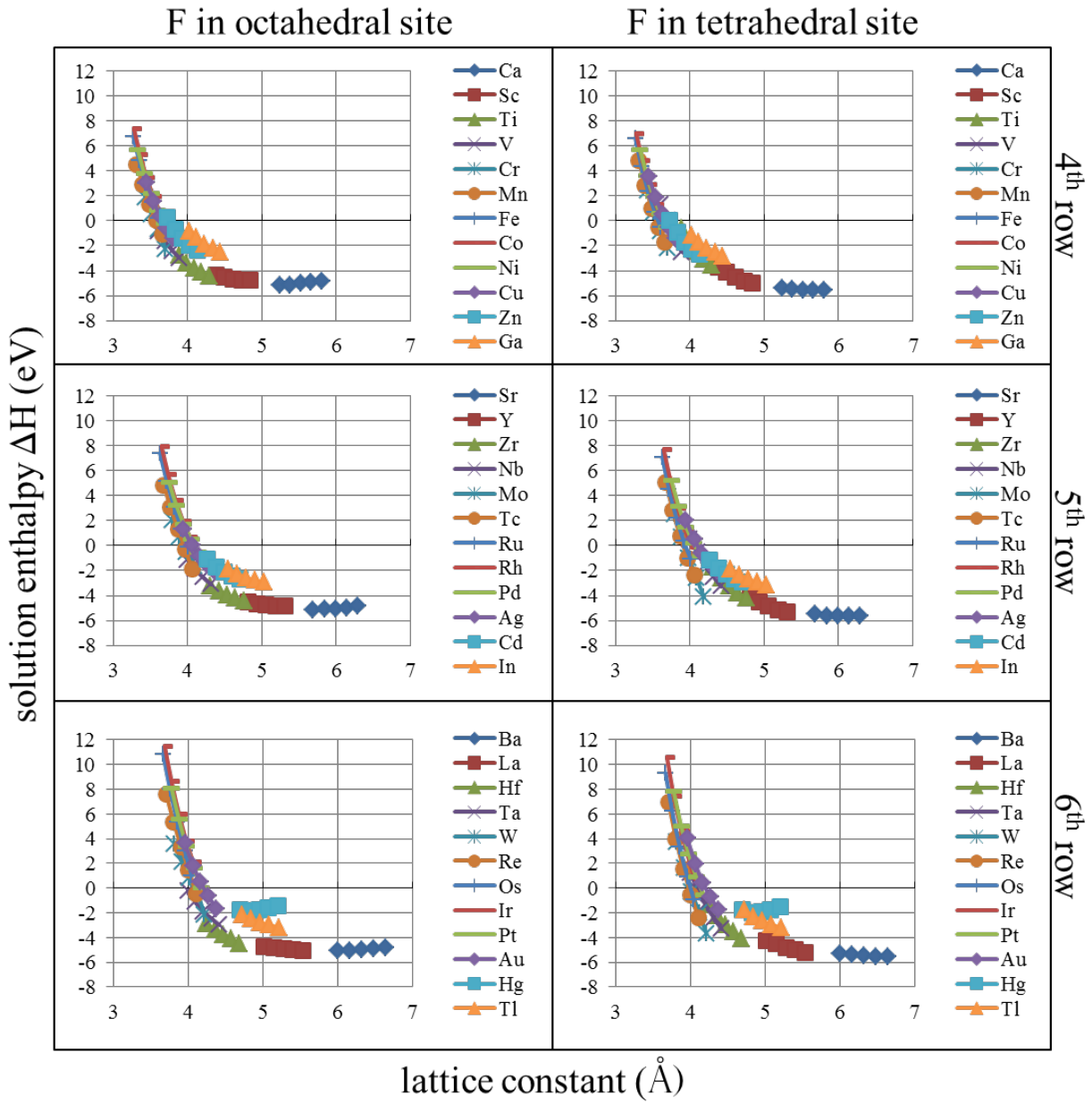


Figure B.7: Solution enthalpy of fluorine in 4th, 5th, and 6th row elements in the octahedral (left) and in the tetrahedral (right) site.

B.3.6 Helium solution enthalpy vs lattice constant

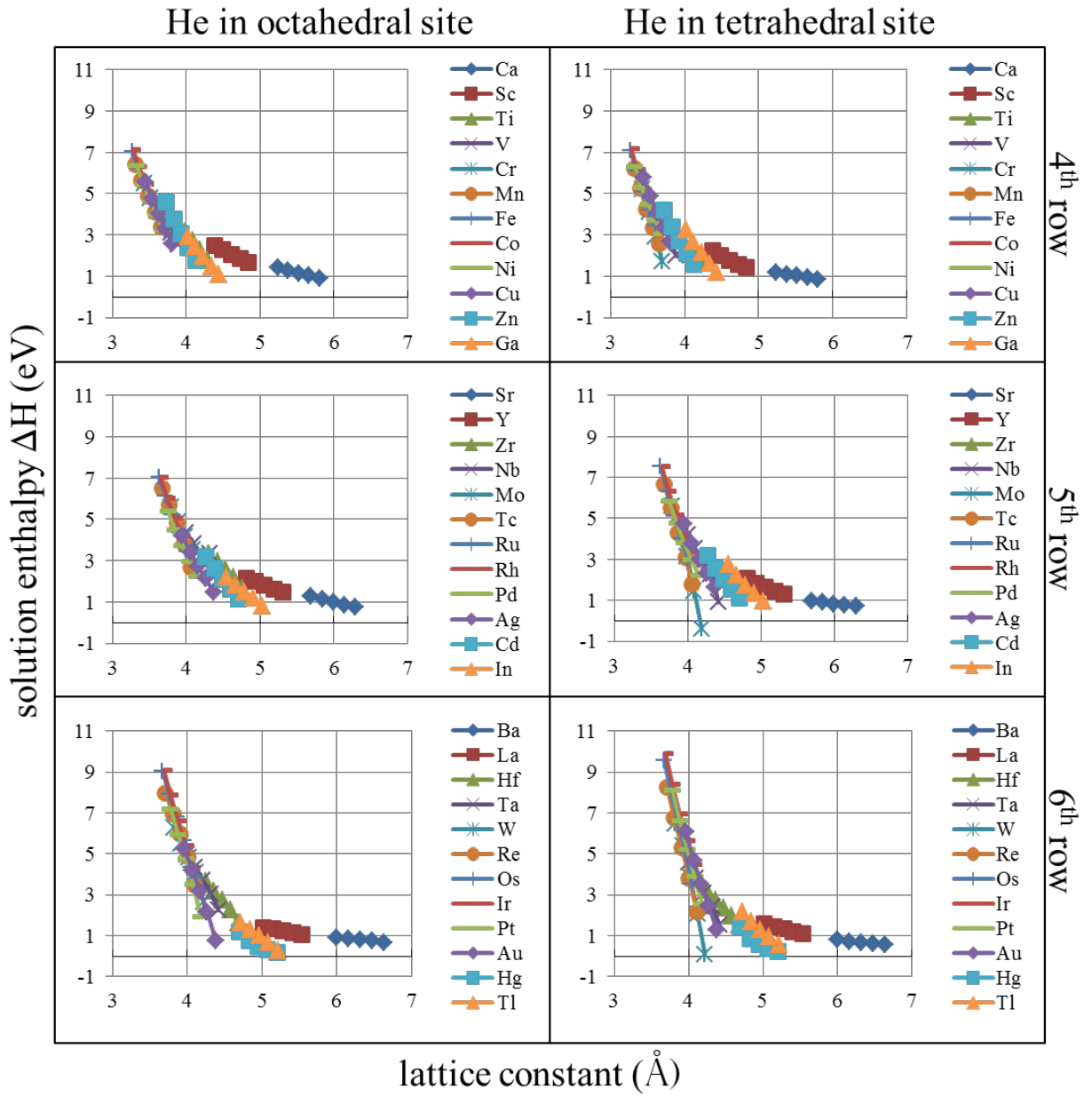


Figure B.8: Solution enthalpy of helium in 4th, 5th, and 6th row elements in the octahedral (left) and in the tetrahedral (right) site.

B.3.7 Neon solution enthalpy vs lattice constant

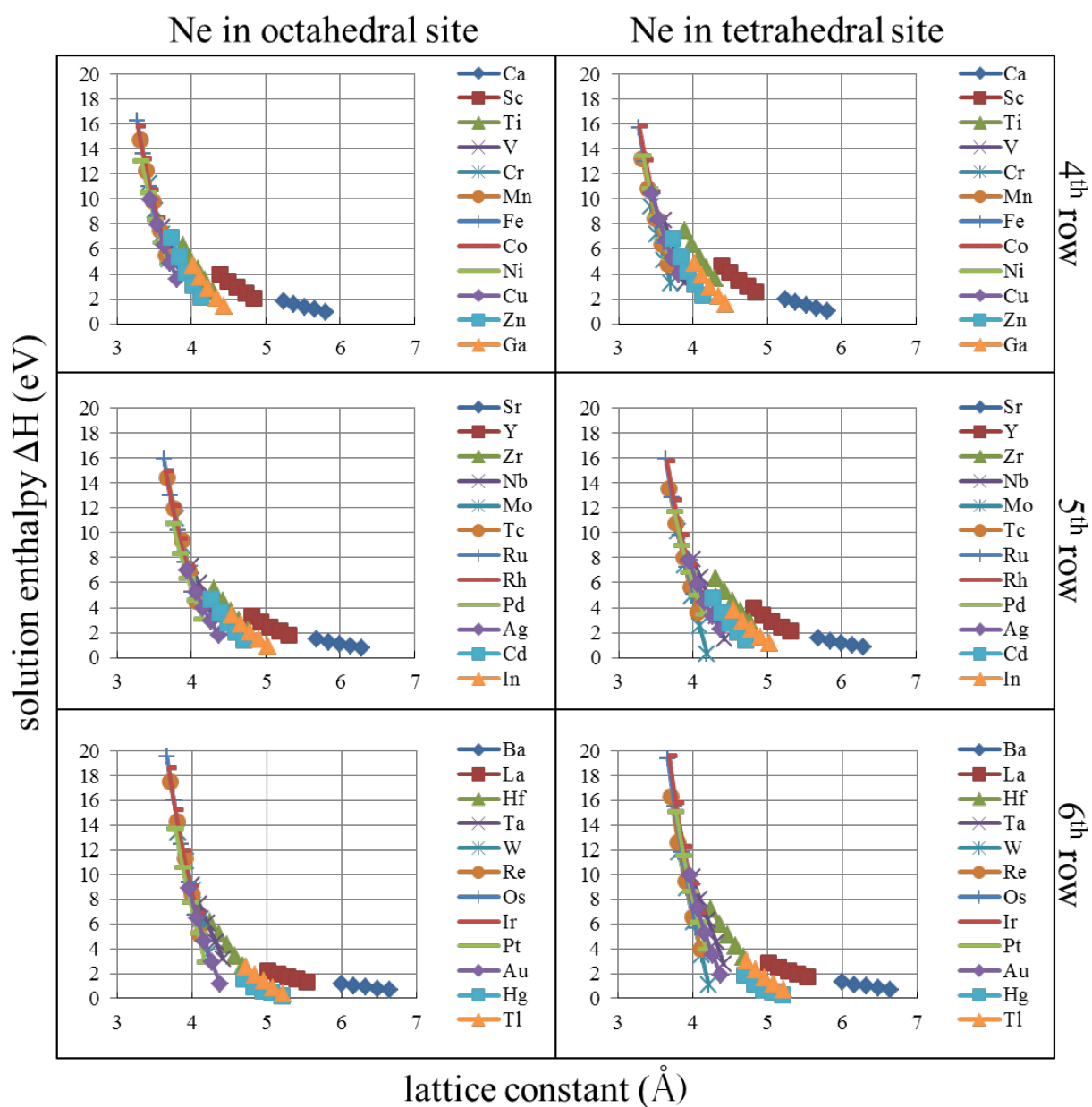


Figure B.9: Solution enthalpy of neon in 4th, 5th, and 6th row elements in the octahedral (left) and in the tetrahedral (right) site.

B.4 Charge density differences of interstitial elements

B.4.1 CDD of Interstitial elements in early TMs

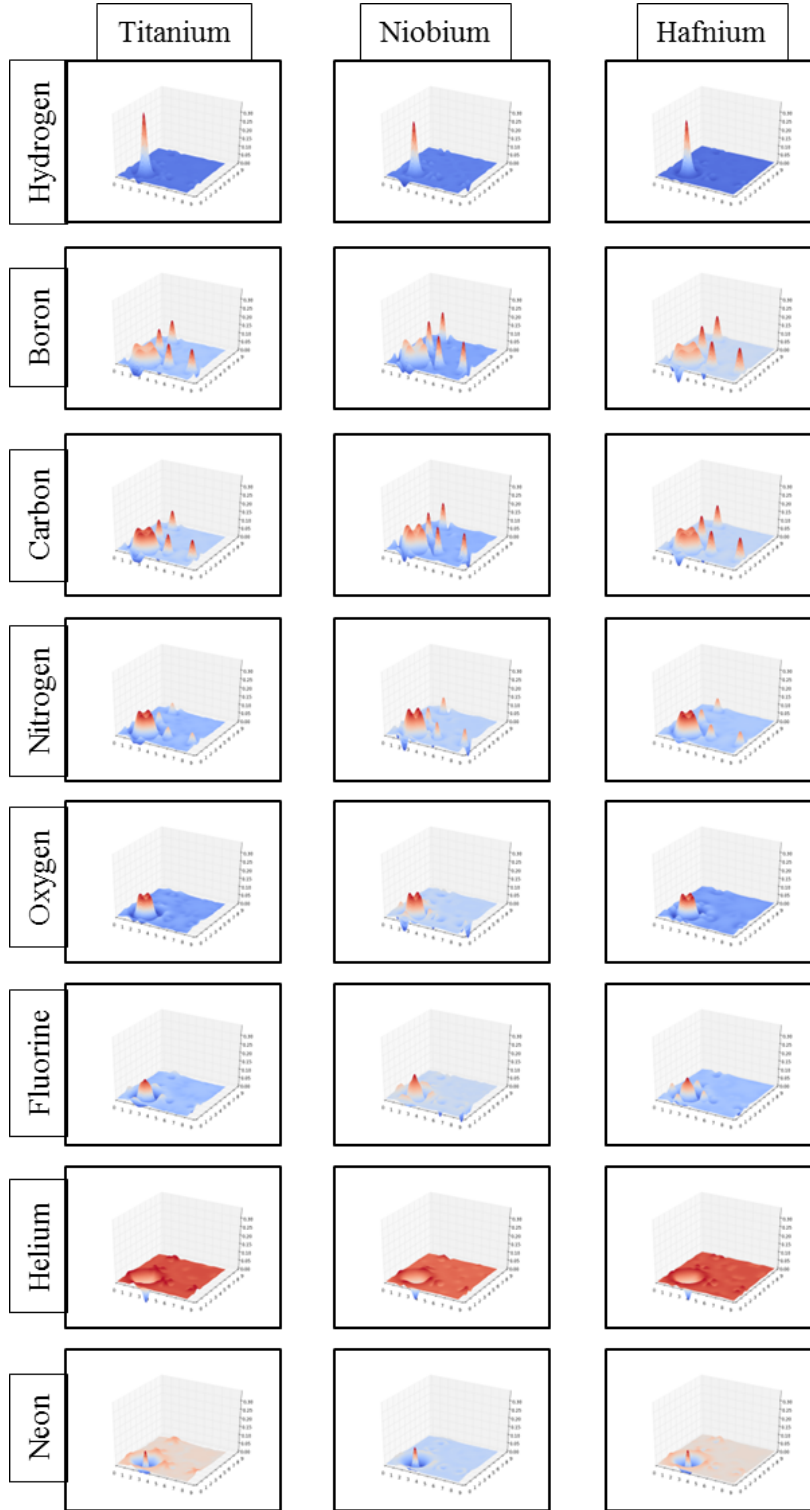


Figure B.10: Charge density difference for early TMs.

B.4.2 CDD of Interstitial elements in half-filled TMs

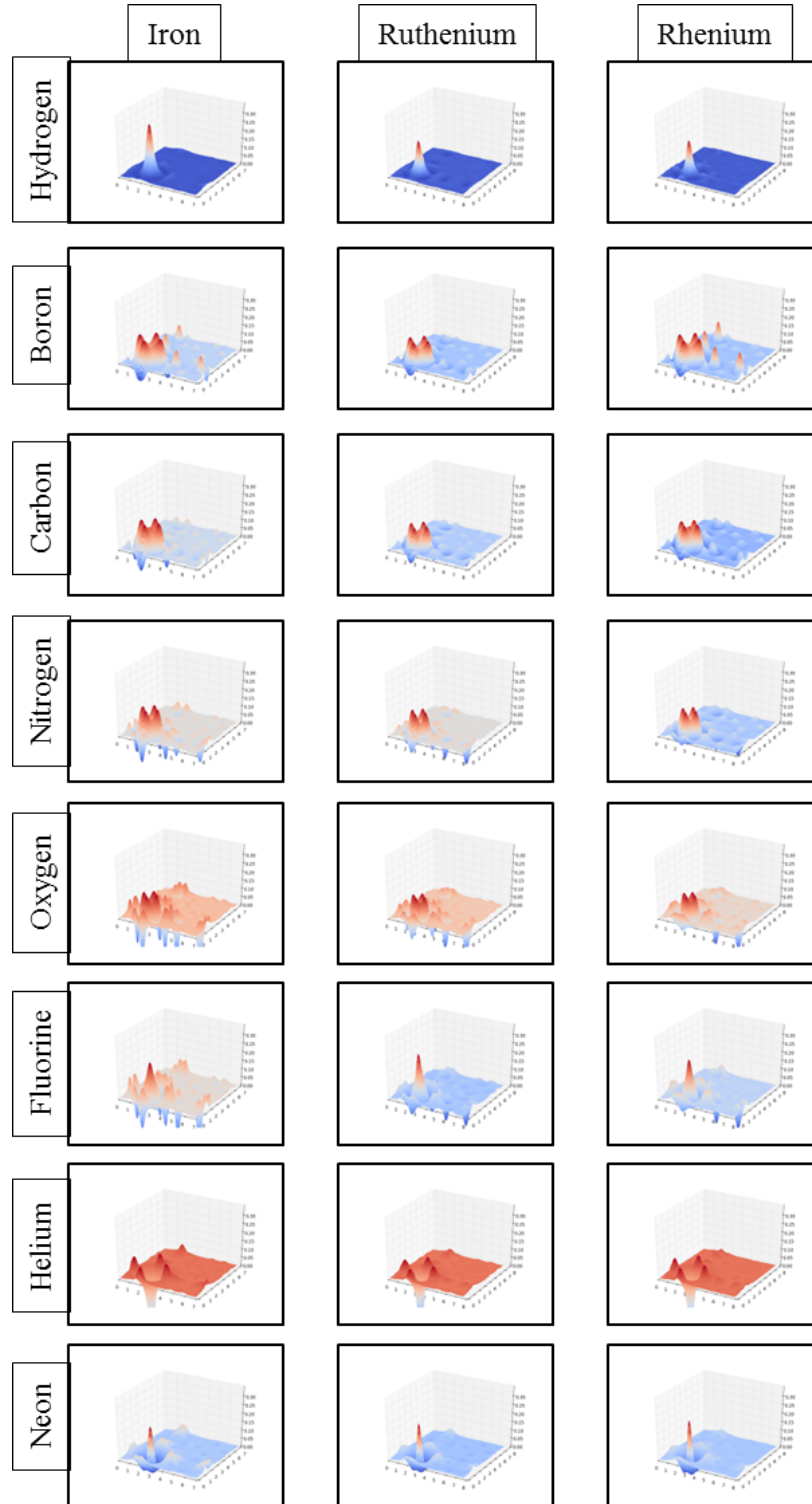


Figure B.11: Charge density difference for half-filled d-band TMs.

B.4.3 CDD of Interstitial elements in late TMs

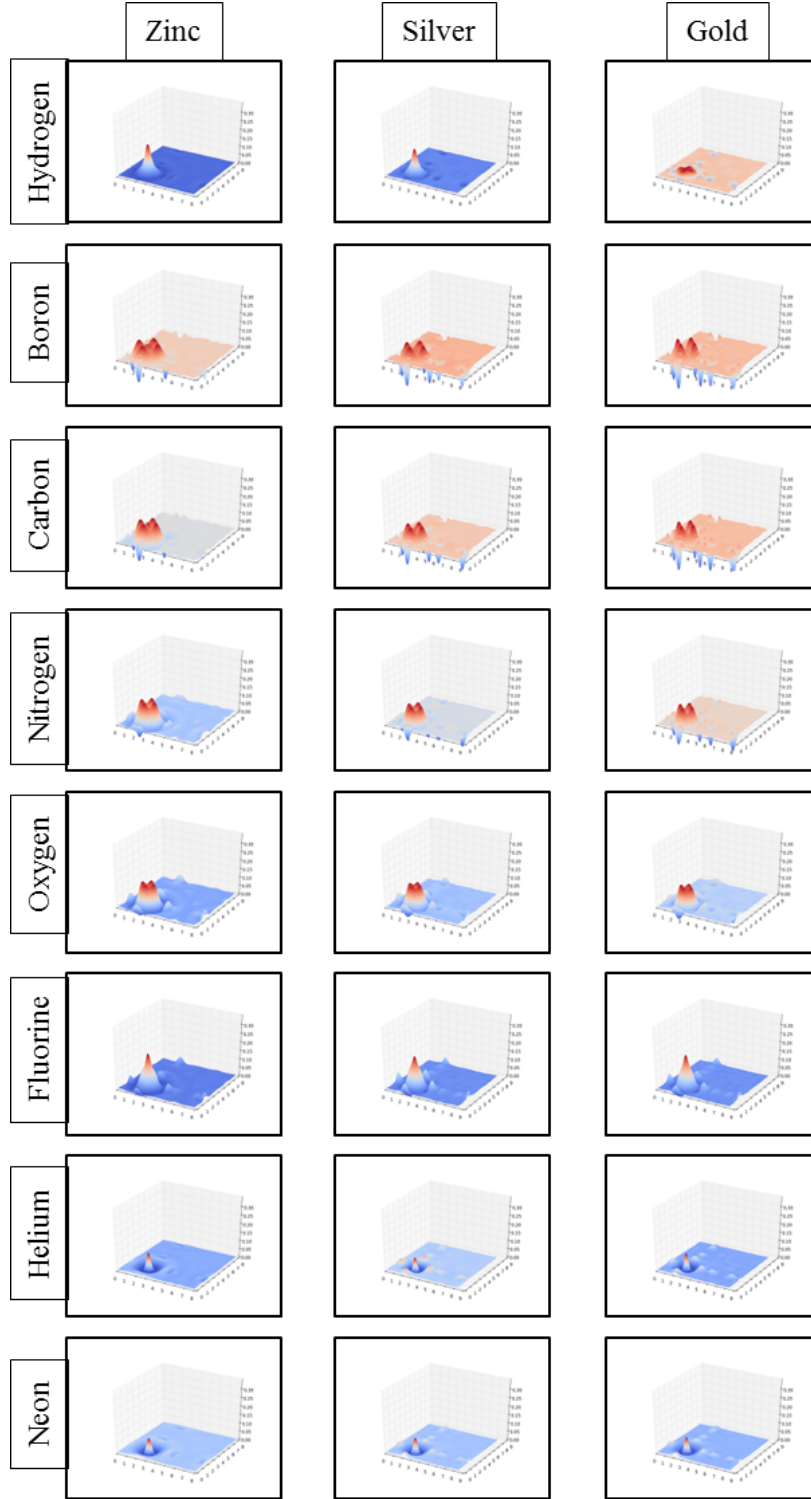


Figure B.12: Charge density difference for filled d-band TMs.

B.4.4 CDD of H in all TMs

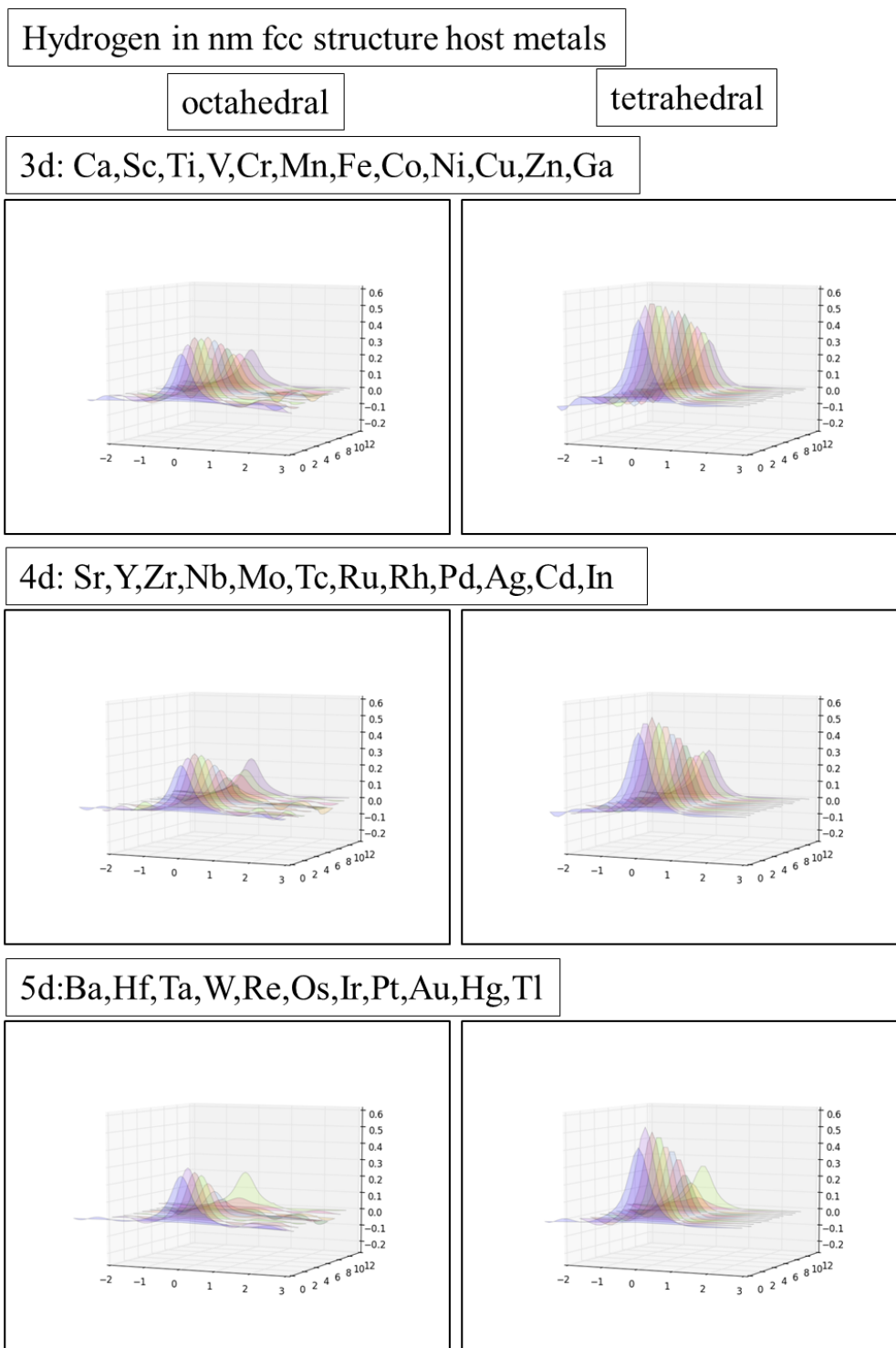


Figure B.13: Charge density difference of hydrogen for all studied elements in octahedral and tetrahedral site.

Appendix C

Factor Analysis

C.1 Solution enthalpy matrix

Pos	Bulk	Hydrogen	Boron	Carbon	Nitrogen	Oxygen	Fluorine	Helium	Neon
0	Ca	-0.687	0.030	0.442	-2.056	-5.912	-5.532	1.016	1.304
1	Sc	-0.943	-0.807	-0.967	-3.411	-6.560	-4.906	1.782	3.120
2	Ti	-0.419	-1.134	-1.550	-3.701	-6.243	-3.716	2.895	4.877
3	V	-0.332	-0.611	-0.840	-2.714	-4.905	-2.144	3.085	6.267
6	Fe	0.167	-0.139	0.554	0.248	-1.241	1.100	4.498	8.592
7	Co	0.045	-0.355	0.716	0.504	-0.746	1.240	4.653	8.334
8	Ni	0.054	-0.911	0.554	0.682	-0.300	1.498	4.484	7.564
9	Cu	0.429	0.569	2.614	2.083	-0.406	0.017	3.802	5.548
10	Zn	1.047	1.266	2.490	0.882	-2.441	-2.010	2.836	3.984
0	Sr	-0.727	0.502	1.152	-1.374	-5.750	-5.665	0.818	1.004
1	Y	-0.959	-0.550	-0.503	-2.928	-6.342	-5.166	1.587	2.494
2	Zr	-0.424	-1.387	-1.674	-3.758	-6.344	-4.036	2.623	4.221
3	Nb	-0.414	-0.808	-0.845	-2.448	-4.502	-1.921	3.137	6.322
4	Mo	0.632	0.496	1.239	0.563	-1.006	1.126	4.929	9.016
5	Tc	-0.163	-1.073	-0.073	-0.327	-1.339	0.845	4.470	8.333
6	Ru	0.261	0.008	1.604	1.882	0.648	3.176	5.302	10.197
7	Rh	0.190	-1.303	0.803	1.571	0.548	2.488	4.958	8.866
8	Pd	-0.129	-2.158	0.363	1.299	0.052	1.095	3.592	5.736
9	Ag	0.674	1.251	4.021	3.264	-0.233	-1.057	2.597	3.524
10	Cd	0.968	1.683	3.177	1.629	-1.915	-2.572	1.715	2.297
2	Hf	-0.352	-1.436	-1.926	-3.983	-6.337	-3.700	3.111	4.925
3	Ta	-0.393	-1.044	-1.245	-2.809	-4.669	-1.835	3.177	6.662
4	W	0.932	0.825	1.555	1.122	0.038	2.255	5.842	10.870
5	Re	0.135	-0.884	0.190	0.279	-0.696	1.818	5.676	10.112
6	Os	0.851	0.872	2.776	3.073	1.825	4.888	6.805	12.735
7	Ir	0.938	-0.291	2.315	2.726	1.662	4.175	6.512	11.322
8	Pt	0.501	-1.390	1.569	1.957	0.943	2.332	4.650	7.033
9	Au	0.822	0.669	4.093	3.709	0.582	0.058	2.936	3.909
11	Tl	0.888	2.080	4.029	2.320	-1.803	-2.935	1.120	1.524

Table C.1: Solution enthalpy of interstitial elements in TM GS (units in eV). Gray shaded elements are not considered in the PCC matrix 6.1.

C.2 Standardized solution enthalpy matrix

	Hydrogen	Boron	Carbon	Nitrogen	Oxygen	Fluorine	Helium	Neon
Sc	-1.844	-0.386	-0.943	-1.339	-1.596	-1.724	-1.608	-1.431
Ti	-0.909	-0.750	-1.282	-1.454	-1.488	-1.305	-0.818	-0.801
V	-0.753	-0.168	-0.869	-1.063	-1.031	-0.751	-0.683	-0.302
Fe	0.138	0.358	-0.059	0.111	0.222	0.391	0.321	0.533
Co	-0.080	0.117	0.035	0.212	0.391	0.441	0.431	0.440
Ni	-0.064	-0.501	-0.059	0.283	0.543	0.531	0.311	0.164
Cu	0.586	1.249	1.147	0.838	0.513	0.063	-0.173	-0.560
Y	-1.873	-0.100	-0.674	-1.148	-1.522	-1.816	-1.747	-1.656
Zr	-0.918	-1.031	-1.354	-1.477	-1.523	-1.418	-1.011	-1.036
Nb	-0.900	-0.387	-0.872	-0.958	-0.893	-0.673	-0.646	-0.282
Mo	0.968	1.065	0.339	0.236	0.302	0.400	0.627	0.685
Tc	-0.452	-0.682	-0.424	-0.117	0.188	0.301	0.301	0.440
Ru	0.306	0.522	0.551	0.758	0.867	1.123	0.892	1.109
Rh	0.175	-0.877	0.115	0.649	0.831	0.854	0.648	0.631
Pd	-0.344	-1.790	-0.096	0.580	0.702	0.421	-0.322	-0.492
Ag	1.081	2.027	1.955	1.297	0.569	-0.329	-1.029	-1.286
Hf	-0.789	-1.087	-1.500	-1.566	-1.520	-1.299	-0.664	-0.783
Ta	-0.862	-0.650	-1.105	-1.101	-0.950	-0.642	-0.617	-0.160
W	1.504	1.432	0.522	0.457	0.659	0.798	1.276	1.351
Re	0.081	-0.472	-0.271	0.123	0.408	0.644	1.158	1.079
Os	1.359	1.484	1.232	1.230	1.269	1.726	1.960	2.020
Ir	1.520	0.238	0.994	1.109	1.198	1.418	1.752	1.513
Pt	0.734	-1.014	0.554	0.808	0.970	0.836	0.429	-0.027
Au	1.338	1.402	2.065	1.531	0.892	0.008	-0.789	-1.148

Table C.2: Standardized solution enthalpy matrix **Z**

C.3 Factor Analysis of ΔH for interstitial elements H, B, C, N, O, F, He, Ne

C.3.1 Determination of the number of factors k

The correlation coefficient matrix can be calculated from the standardized solution enthalpy table C.2

$$\mathbf{R} = \frac{1}{n-1} \mathbf{Z}^\dagger \mathbf{Z} = \begin{pmatrix} 1.00 & 0.65 & 0.85 & 0.88 & 0.86 & 0.80 & 0.71 & 0.59 \\ 0.65 & 1.00 & 0.75 & 0.56 & 0.42 & 0.31 & 0.25 & 0.21 \\ 0.85 & 0.75 & 1.00 & 0.94 & 0.83 & 0.64 & 0.41 & 0.28 \\ 0.88 & 0.56 & 0.94 & 1.00 & 0.96 & 0.83 & 0.60 & 0.47 \\ 0.86 & 0.42 & 0.83 & 0.96 & 1.00 & 0.94 & 0.76 & 0.65 \\ 0.80 & 0.31 & 0.64 & 0.83 & 0.94 & 1.00 & 0.92 & 0.86 \\ 0.71 & 0.25 & 0.41 & 0.60 & 0.76 & 0.92 & 1.00 & 0.97 \\ 0.59 & 0.21 & 0.28 & 0.47 & 0.65 & 0.86 & 0.97 & 1.00 \end{pmatrix} \quad (\text{C.1})$$

Using the relation

$$\mathbf{R} = \mathbf{A} \mathbf{A}^\dagger. \quad (\text{C.2})$$

and by using the singular value decomposition (svd) of the symmetric matrix \mathbf{R} reveals

$$\mathbf{R} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^\dagger = \mathbf{U} \mathbf{\Sigma} \mathbf{U}^\dagger. \quad (\text{C.3})$$

in which $\mathbf{\Sigma}$ is a diagonal matrix with the eigenvalues of \mathbf{R} , thus

$$\mathbf{\Sigma} = \text{diag} \left(5.3876 \quad 1.4730 \quad 0.5436 \quad 0.1184 \quad 0.0170 \quad 0.0071 \quad 0.0025 \quad 0.0008. \right) \quad (\text{C.4})$$

and column-wise the corresponding eigenvectors

$$\mathbf{U} = \begin{pmatrix} -0.388 & -0.136 & 0.087 & -0.877 & -0.219 & -0.071 & -0.044 & -0.024 \\ -0.241 & -0.502 & 0.715 & 0.277 & -0.176 & 0.250 & 0.064 & -0.065 \\ -0.347 & -0.433 & -0.150 & 0.103 & 0.494 & -0.494 & 0.183 & 0.371 \\ -0.384 & -0.207 & -0.356 & 0.179 & 0.099 & 0.030 & -0.412 & -0.685 \\ -0.398 & 0.007 & -0.358 & 0.175 & -0.304 & 0.456 & -0.237 & 0.571 \\ -0.389 & 0.256 & -0.153 & 0.193 & -0.378 & -0.142 & 0.715 & -0.220 \\ -0.345 & 0.429 & 0.206 & -0.097 & 0.645 & 0.453 & 0.147 & -0.043 \\ -0.307 & 0.500 & 0.371 & 0.186 & -0.123 & -0.505 & -0.449 & 0.107 \end{pmatrix} \quad (\text{C.5})$$

Comparing equations C.2 and C.3

$$\mathbf{R} = \mathbf{A} \mathbf{A}^\dagger = \mathbf{U} \mathbf{\Sigma} \mathbf{U}^\dagger \quad (\text{C.6})$$

matrix \mathbf{A} is equal to $\mathbf{U} \mathbf{\Sigma}^{1/2} \mathbf{T}$, in which \mathbf{T} corresponds to any orthogonal rotation matrix. For $\mathbf{T} = \mathbf{I}$ (identity) it follows for the loading matrix

$$\mathbf{A} = \begin{pmatrix} 0.936 & 0.165 & -0.064 & 0.302 & 0.029 & 0.006 & 0.002 & 0.001 \\ 0.583 & 0.610 & -0.527 & -0.096 & 0.023 & -0.021 & -0.003 & 0.002 \\ 0.839 & 0.525 & 0.111 & -0.036 & -0.064 & 0.042 & -0.009 & -0.010 \\ 0.929 & 0.252 & 0.263 & -0.062 & -0.013 & -0.003 & 0.021 & 0.019 \\ 0.961 & -0.009 & 0.264 & -0.060 & 0.040 & -0.038 & 0.012 & -0.016 \\ 0.939 & -0.311 & 0.113 & -0.066 & 0.049 & 0.012 & -0.036 & 0.006 \\ 0.835 & -0.520 & -0.152 & 0.033 & -0.084 & -0.038 & -0.007 & 0.001 \\ 0.742 & -0.607 & -0.274 & -0.064 & 0.016 & 0.043 & 0.023 & -0.003 \end{pmatrix}. \quad (\text{C.7})$$

The reduction of the dimension to k factors can be achieved with different methods. One method to reduce the dimension of the problem is the usage of the Kaiser-Guttman-Criteria: All factors with eigenvalues larger than one will mainly determine the originally formulated standardized variables. In the formulated problem above there are only two eigenvalues which are larger than

one (i.e. $k = 2$). The reduced loading matrix is then defined as the first two columns of A , i.e.

$$\mathbf{A}_{\text{red}^2} = \begin{pmatrix} 0.936 & 0.165 \\ 0.583 & 0.610 \\ 0.839 & 0.525 \\ 0.929 & 0.252 \\ 0.961 & -0.009 \\ 0.939 & -0.311 \\ 0.835 & -0.520 \\ 0.742 & -0.607 \end{pmatrix} \mathbf{T}, \quad (\text{C.8})$$

in which for $k = 2$ the orthogonal transformation matrix corresponds to

$$\mathbf{T} = \begin{pmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{pmatrix} \quad (\text{C.9})$$

According to equation C.2 the reduced correlation coefficient matrix can be achieved (independent of \mathbf{T}) through $\mathbf{R}_{\text{red}^2} = \mathbf{A}_{\text{red}^2} \mathbf{A}_{\text{red}^2}^\dagger$

$$\mathbf{R}_{\text{red}^2} = \begin{pmatrix} 0.9 & 0.65 & 0.87 & 0.91 & 0.9 & 0.83 & 0.7 & 0.59 \\ 0.65 & 0.71 & 0.81 & 0.7 & 0.56 & 0.36 & 0.17 & 0.06 \\ 0.87 & 0.81 & 0.98 & 0.91 & 0.8 & 0.63 & 0.43 & 0.3 \\ 0.91 & 0.7 & 0.91 & 0.93 & 0.89 & 0.79 & 0.64 & 0.54 \\ 0.9 & 0.56 & 0.8 & 0.89 & 0.92 & 0.91 & 0.81 & 0.72 \\ 0.83 & 0.36 & 0.63 & 0.79 & 0.91 & 0.98 & 0.95 & 0.89 \\ 0.7 & 0.17 & 0.43 & 0.64 & 0.81 & 0.95 & 0.97 & 0.93 \\ 0.59 & 0.06 & 0.3 & 0.54 & 0.72 & 0.89 & 0.93 & 0.92 \end{pmatrix}. \quad (\text{C.10})$$

However, the reduction to three factors captures all correlation coefficients very well. The three-components reduced PCC is given as follows

$$\mathbf{R}_{\text{red}^3} = \begin{pmatrix} 0.91 & 0.68 & 0.87 & 0.89 & 0.88 & 0.82 & 0.71 & 0.61 \\ 0.68 & 0.99 & 0.75 & 0.56 & 0.42 & 0.30 & 0.25 & 0.21 \\ 0.87 & 0.75 & 0.99 & 0.94 & 0.83 & 0.64 & 0.41 & 0.27 \\ 0.89 & 0.56 & 0.94 & 1.00 & 0.96 & 0.82 & 0.60 & 0.46 \\ 0.88 & 0.42 & 0.83 & 0.96 & 0.99 & 0.94 & 0.77 & 0.65 \\ 0.82 & 0.30 & 0.64 & 0.82 & 0.94 & 0.99 & 0.93 & 0.85 \\ 0.71 & 0.25 & 0.41 & 0.60 & 0.77 & 0.93 & 0.99 & 0.98 \\ 0.61 & 0.21 & 0.27 & 0.46 & 0.65 & 0.85 & 0.98 & 0.99 \end{pmatrix}. \quad (\text{C.11})$$

C.4 Results/Tables for the combinatoric search

Formula	ΔH function	correlation
$\Gamma_{(0,0,0)}$	$0.00900 + 0.00900 + 0.00900$	0.000
$\Gamma_{(0,0,1)}$	$(-0.60800) + (-0.60800) + 1.24500 \cdot \widetilde{\Delta E_{F-s}}$	0.879
$\Gamma_{(0,0,2)}$	$(-0.49100) + (-0.49100) + 0.97100 \cdot \widetilde{R_j} \cdot \widetilde{\Delta E_{F-s}}$	0.884
$\Gamma_{(0,0,3)}$	$(-0.53100) + (-0.53100) + 1.07700 \cdot \widetilde{N_{tot}^{-1}} \cdot \widetilde{R_j} \cdot \widetilde{\Delta E_{F-s}}$	0.906
$\Gamma_{(0,1,1)}$	$(-1.51900) + 1.12400 \cdot \widetilde{\Delta E_{F-s}} + 0.42400 \cdot \widetilde{R_j}$	0.910
$\Gamma_{(0,1,2)}$	$(-1.69500) + 0.58000 \cdot \widetilde{R_j} + 1.15900 \cdot \widetilde{N_{tot}^{-1}} \cdot \widetilde{\Delta E_{F-s}}$	0.915
$\Gamma_{(0,1,3)}$	$(-1.69500) + 0.58000 \cdot \widetilde{R_j} + 1.15900 \cdot \widetilde{N_{tot}^{-1}} \cdot 1 \cdot \widetilde{\Delta E_{F-s}}$	0.915
$\Gamma_{(0,2,2)}$	$(-1.50300) + 0.28600 \cdot \widetilde{R_j} \cdot \widetilde{R_j} + 1.23200 \cdot \widetilde{N_{tot}^{-1}} \cdot \widetilde{\Delta E_{F-s}}$	0.918
$\Gamma_{(0,2,3)}$	$(-1.23800) + 1.90300 \cdot \widetilde{R_j} \cdot \widetilde{\Delta E_{F-s}} + (-0.58300) \cdot \widetilde{R_j} \cdot \widetilde{B_0} \cdot \widetilde{f_d}$	0.927
$\Gamma_{(0,3,3)}$	$(-0.16400) + (-0.98600) \cdot \widetilde{R_j^{-1}} \cdot \widetilde{f_d^{-1}} \cdot \widetilde{B_0} + 1.20500 \cdot \widetilde{f_d^{-1}} \cdot \widetilde{\Delta E_{F-s}} \cdot \widetilde{\Delta E_{F-s}}$	0.940
$\Gamma_{(1,1,1)}$	$1.12400 \cdot \widetilde{\Delta E_{F-s}} + 0.42400 \cdot \widetilde{R_j} + (-1.51900) \cdot 1$	0.910
$\Gamma_{(1,1,2)}$	$(-1.69500) \cdot 1 + 0.58000 \cdot \widetilde{R_j} + 1.15900 \cdot \widetilde{N_{tot}^{-1}} \cdot \widetilde{\Delta E_{F-s}}$	0.915
$\Gamma_{(1,1,3)}$	$2.55700 \cdot \widetilde{\Delta E_{F-s}} + (-1.59300) \cdot \widetilde{N_{tot}^{-1}} + (-1.03500) \cdot \widetilde{R_j^{-1}} \cdot \widetilde{B_0} \cdot \widetilde{f_d}$	0.930
$\Gamma_{(1,2,2)}$	$2.30500 \cdot \widetilde{\Delta E_{F-s}} + (-1.13300) \cdot \widetilde{f_d^{-1}} \cdot \widetilde{B_0} + (-1.63000) \cdot \widetilde{R_j^{-1}} \cdot \widetilde{f_d}$	0.921
$\Gamma_{(1,2,3)}$	$(-2.20200) \cdot \widetilde{W_d^{1/2}} + 3.25400 \cdot \widetilde{N_{tot}^{-1}} \cdot \widetilde{\Delta E_{F-s}} + (-1.08100) \cdot \widetilde{R_j^{-1}} \cdot \widetilde{B_0} \cdot \widetilde{f_d}$	0.949
$\Gamma_{(1,3,3)}$	$(-0.46500) \cdot \widetilde{f_d^{-1}} + 1.92200 \cdot \widetilde{f_d^{-1}} \cdot \widetilde{\Delta E_{F-s}} \cdot \widetilde{\Delta E_{F-s}} + (-1.14500) \cdot \widetilde{R_j^{-1}} \cdot \widetilde{W_d^{1/2}} \cdot \widetilde{B_0}$	0.951
$\Gamma_{(2,2,2)}$	$(-0.11100) \cdot \widetilde{R_j} \cdot \widetilde{B_0} + 0.35700 \cdot \widetilde{R_j} \cdot \widetilde{\Delta E_{F-s}} + (-0.03900) \cdot \widetilde{f_d} \cdot \widetilde{f_d}$	0.929
$\Gamma_{(2,2,3)}$	$(-2.20200) \cdot 1 \cdot \widetilde{W_d^{1/2}} + 3.25400 \cdot \widetilde{N_{tot}^{-1}} \cdot \widetilde{\Delta E_{F-s}} + (-1.08100) \cdot \widetilde{R_j^{-1}} \cdot \widetilde{B_0} \cdot \widetilde{f_d}$	0.949
$\Gamma_{(2,3,3)}$	$(-0.67400) \cdot \widetilde{R_j^{-1}} \cdot \widetilde{B_0} + 1.77500 \cdot \widetilde{f_d^{-1}} \cdot \widetilde{\Delta E_{F-s}} \cdot \widetilde{\Delta E_{F-s}} + (-1.02300) \cdot \widetilde{R_j^{-1}} \cdot \widetilde{f_d} \cdot \widetilde{B_0}$	0.954
$\Gamma_{(3,3,3)}$	$(-0.94300) \cdot \widetilde{R_j^{-1}} \cdot \widetilde{\Delta E_{F-s}} \cdot \widetilde{B_0} + 3.35200 \cdot \widetilde{f_d^{-1}} \cdot \widetilde{\Delta E_{F-s}} \cdot \widetilde{\Delta E_{F-s}} + (-2.09600) \cdot \widetilde{f_d^{-1}} \cdot \widetilde{W_d^{1/2}} \cdot \widetilde{\Delta E_{F-s}}$	0.964

Table C.3: Results of the combinatorial search for the best fitting formula for hydrogen

Formula	ΔH function	correlation
$\Gamma_{(0,0,0)}$	$0.11400 + 0.11400 + 0.11400$	0.000
$\Gamma_{(0,0,1)}$	$(-1.30100) + (-1.30100) + 2.94400 \cdot \widetilde{\Delta E_{F-s}}$	0.799
$\Gamma_{(0,0,2)}$	$(-1.13800) + (-1.13800) + 2.51700 \cdot \widetilde{R_j} \cdot \widetilde{\Delta E_{F-s}}$	0.881
$\Gamma_{(0,0,3)}$	$(-1.06500) + (-1.06500) + 2.31800 \cdot \widetilde{R_j} \cdot \widetilde{\Delta E_{F-s}} \cdot \widetilde{N_{tot}}$	0.881
$\Gamma_{(0,1,1)}$	$(-2.79800) + 1.17000 \cdot \widetilde{f_d} + 1.97000 \cdot \widetilde{\Delta E_{F-s}}$	0.855
$\Gamma_{(0,1,2)}$	$(-4.44200) + 0.94100 \cdot \widetilde{B_0^{-1}} + 3.69400 \cdot \widetilde{R_j} \cdot \widetilde{\Delta E_{F-s}}$	0.946
$\Gamma_{(0,1,3)}$	$(-5.09700) + 1.12500 \cdot \widetilde{B_0^{-1}} + 4.25800 \cdot \widetilde{N_{tot}^{-1}} \cdot \widetilde{R_j} \cdot \widetilde{\Delta E_{F-s}}$	0.949
$\Gamma_{(0,2,2)}$	$(-4.49400) + 3.78100 \cdot \widetilde{R_j} \cdot \widetilde{\Delta E_{F-s}} + 0.87900 \cdot \widetilde{B_0^{-1}} \cdot \widetilde{N_{tot}^{-1}}$	0.947
$\Gamma_{(0,2,3)}$	$(-5.14100) + 1.25500 \cdot \widetilde{B_0^{-1}} \cdot \widetilde{N_{tot}} + 4.20600 \cdot \widetilde{N_{tot}^{-1}} \cdot \widetilde{R_j} \cdot \widetilde{\Delta E_{F-s}}$	0.953
$\Gamma_{(0,3,3)}$	$(-5.04800) + 1.33000 \cdot \widetilde{B_0^{-1}} \cdot \widetilde{N_{tot}} \cdot \widetilde{N_{tot}^{-1}} \cdot \widetilde{R_j} \cdot \widetilde{\Delta E_{F-s}}$	0.954
$\Gamma_{(1,1,1)}$	$0.55600 \cdot \widetilde{B_0^{-1}} + (-4.31700) \cdot \widetilde{W_d^{1/2}} + 4.13500 \cdot \widetilde{\Delta E_{F-s}}$	0.882
$\Gamma_{(1,1,2)}$	$(-4.44200) \cdot 1 + 0.94100 \cdot \widetilde{B_0^{-1}} + 3.69400 \cdot \widetilde{R_j} \cdot \widetilde{\Delta E_{F-s}}$	0.946
$\Gamma_{(1,1,3)}$	$1.12500 \cdot \widetilde{B_0^{-1}} + (-5.09700) \cdot 1 + 4.25800 \cdot \widetilde{N_{tot}^{-1}} \cdot \widetilde{R_j} \cdot \widetilde{\Delta E_{F-s}}$	0.949
$\Gamma_{(1,2,2)}$	$(-5.24700) \cdot \widetilde{N_{tot}} + 4.33700 \cdot \widetilde{R_j} \cdot \widetilde{\Delta E_{F-s}} + 1.08000 \cdot \widetilde{W_d^{-1/2}} \cdot \widetilde{B_0^{-1}}$	0.950
$\Gamma_{(1,2,3)}$	$(-5.24000) \cdot \widetilde{N_{tot}} + 4.21600 \cdot \widetilde{R_j} \cdot \widetilde{\Delta E_{F-s}} + 1.25500 \cdot \widetilde{B_0^{-1}} \cdot \widetilde{N_{tot}} \cdot \widetilde{N_{tot}^{-1}}$	0.954
$\Gamma_{(1,3,3)}$	$(-5.04800) \cdot 1 + 1.33000 \cdot \widetilde{B_0^{-1}} \cdot \widetilde{N_{tot}} \cdot \widetilde{N_{tot}^{-1}} \cdot \widetilde{R_j} \cdot \widetilde{\Delta E_{F-s}}$	0.954
$\Gamma_{(2,2,2)}$	$(-5.24700) \cdot 1 \cdot \widetilde{N_{tot}} + 4.33700 \cdot \widetilde{R_j} \cdot \widetilde{\Delta E_{F-s}} + 1.08000 \cdot \widetilde{W_d^{-1/2}} \cdot \widetilde{B_0^{-1}}$	0.950
$\Gamma_{(2,2,3)}$	$4.21600 \cdot \widetilde{R_j} \cdot \widetilde{\Delta E_{F-s}} + (-5.24000) \cdot 1 \cdot \widetilde{N_{tot}} + 1.25500 \cdot \widetilde{B_0^{-1}} \cdot \widetilde{N_{tot}} \cdot \widetilde{N_{tot}^{-1}}$	0.954
$\Gamma_{(2,3,3)}$	$(-3.95600) \cdot \widetilde{W_d^{1/2}} \cdot \widetilde{N_{tot}} + 3.51600 \cdot \widetilde{W_d^{1/2}} \cdot \widetilde{R_j} \cdot \widetilde{\Delta E_{F-s}} + 0.75200 \cdot \widetilde{B_0^{-1}} \cdot \widetilde{f_d}$	0.956
$\Gamma_{(3,3,3)}$	$0.75200 \cdot \widetilde{B_0^{-1}} \cdot \widetilde{B_0^{-1}} \cdot \widetilde{f_d} + (-3.95600) \cdot 1 \cdot \widetilde{W_d^{1/2}} \cdot \widetilde{N_{tot}} + 3.51600 \cdot \widetilde{W_d^{1/2}} \cdot \widetilde{R_j} \cdot \widetilde{\Delta E_{F-s}}$	0.956

Table C.4: Results of the combinatorial search for the best fitting formula for carbon

Formula	ΔH function	correlation
$\Gamma_{(0,0,0)}$	$(-0.11900) + (-0.11900) + (-0.11900)$	0.000
$\Gamma_{(0,0,1)}$	$(-2.27900) + (-2.27900) + 4.20200 \cdot \tilde{f}_d$	0.864
$\Gamma_{(0,0,2)}$	$(-3.43600) + (-3.43600) + 6.65900 \cdot \widetilde{W_d^{-1/2}} \cdot \widetilde{\Delta E_{F-s}}$	0.946
$\Gamma_{(0,0,3)}$	$(-3.43600) + (-3.43600) + 6.65900 \cdot \widetilde{W_d^{-1/2}} \cdot 1 \cdot \widetilde{\Delta E_{F-s}}$	0.946
$\Gamma_{(0,1,1)}$	$(-6.11100) + 2.56500 \cdot \tilde{f}_d + 3.19000 \cdot \widetilde{\Delta E_{F-s}}$	0.948
$\Gamma_{(0,1,2)}$	$(-9.26200) + 0.86200 \cdot \widetilde{B_0^{-1}} + 8.22000 \cdot \widetilde{W_d^{-1/2}} \cdot \widetilde{\Delta E_{F-s}}$	0.965
$\Gamma_{(0,1,3)}$	$(-11.38500) + 8.04100 \cdot \widetilde{\Delta E_{F-s}} + 3.69100 \cdot \widetilde{\Delta E_{F-s}^{-1}} \cdot \widetilde{\Delta E_{F-s}^{-1}} \cdot \tilde{f}_d$	0.966
$\Gamma_{(0,2,2)}$	$(-13.54200) + 10.58000 \cdot \widetilde{W_d^{-1/2}} \cdot \widetilde{\Delta E_{F-s}} + 2.87000 \cdot \widetilde{\Delta E_{F-s}^{-1}} \cdot \widetilde{W_d^{-1/2}}$	0.970
$\Gamma_{(0,2,3)}$	$(-13.11100) + 10.54200 \cdot \widetilde{W_d^{-1/2}} \cdot \widetilde{\Delta E_{F-s}} + 2.40200 \cdot \widetilde{\Delta E_{F-s}^{-1}} \cdot \widetilde{N_{tot}^{-1}} \cdot \widetilde{W_d^{-1/2}}$	0.971
$\Gamma_{(0,3,3)}$	$(-10.56100) + 9.23400 \cdot \widetilde{W_d^{-1/2}} \cdot \widetilde{N_{tot}^{-1}} \cdot \widetilde{\Delta E_{F-s}} + 1.72200 \cdot \widetilde{\Delta E_{F-s}^{-1}} \cdot \widetilde{B_0^{-1}} \cdot \tilde{f}_d$	0.974
$\Gamma_{(1,1,1)}$	$2.56500 \cdot \tilde{f}_d + 3.19000 \cdot \widetilde{\Delta E_{F-s}} + (-6.11100) \cdot 1$	0.948
$\Gamma_{(1,1,2)}$	$3.43500 \cdot \widetilde{\Delta E_{F-s}} + (-6.78400) \cdot \tilde{f}_d^{-1} + 2.33100 \cdot \widetilde{\Delta E_{F-s}^{-1}} \cdot \tilde{f}_d^{-1}$	0.971
$\Gamma_{(1,1,3)}$	$(-4.88200) \cdot \tilde{f}_d^{-1} + 3.15500 \cdot \widetilde{\Delta E_{F-s}} + 0.68000 \cdot \widetilde{\Delta E_{F-s}^{-1}} \cdot \widetilde{\Delta E_{F-s}^{-1}} \cdot \tilde{f}_d^{-1}$	0.976
$\Gamma_{(1,2,2)}$	$(-7.45700) \cdot \tilde{f}_d^{-1} + 3.78500 \cdot \widetilde{N_{tot}^{-1}} \cdot \widetilde{\Delta E_{F-s}} + 2.60000 \cdot \widetilde{\Delta E_{F-s}^{-1}} \cdot \tilde{f}_d^{-1}$	0.973
$\Gamma_{(1,2,3)}$	$(-5.38900) \cdot \tilde{f}_d^{-1} + 3.33600 \cdot \widetilde{N_{tot}^{-1}} \cdot \widetilde{\Delta E_{F-s}} + 1.21400 \cdot \widetilde{W_d^{-1/2}} \cdot \widetilde{B_0^{-1}} \cdot \tilde{f}_d^{-1}$	0.977
$\Gamma_{(1,3,3)}$	$(-3.67900) \cdot \tilde{f}_d^{-1} + 0.49700 \cdot \widetilde{\Delta E_{F-s}^{-1}} \cdot \widetilde{\Delta E_{F-s}^{-1}} \cdot \tilde{f}_d^{-1} + 2.11700 \cdot \widetilde{W_d^{-1/2}} \cdot \widetilde{\Delta E_{F-s}} \cdot \widetilde{\Delta E_{F-s}}$	0.978
$\Gamma_{(2,2,2)}$	$2.60000 \cdot \widetilde{\Delta E_{F-s}^{-1}} \cdot \widetilde{f}_d^{-1} + (-7.45700) \cdot \tilde{f}_d^{-1} \cdot 1 + 3.78500 \cdot \widetilde{N_{tot}^{-1}} \cdot \widetilde{\Delta E_{F-s}}$	0.973
$\Gamma_{(2,2,3)}$	$(-5.43700) \cdot \widetilde{\Delta E_{F-s}^{-1}} \cdot \widetilde{f}_d^{-1} + 2.23300 \cdot \widetilde{\Delta E_{F-s}} \cdot \widetilde{N_{tot}} + 2.17600 \cdot \widetilde{\Delta E_{F-s}^{-1}} \cdot \widetilde{\Delta E_{F-s}^{-1}} \cdot \tilde{f}_d^{-1}$	0.978
$\Gamma_{(2,3,3)}$	$(-4.45300) \cdot \tilde{f}_d^{-1} \cdot \widetilde{W_d^{-1/2}} + 3.21900 \cdot \widetilde{N_{tot}^{-1}} \cdot 1 \cdot \widetilde{\Delta E_{F-s}} + 0.31800 \cdot \widetilde{B_0^{-1}} \cdot \tilde{f}_d^{-1} \cdot \tilde{f}_d^{-1}$	0.981
$\Gamma_{(3,3,3)}$	$0.31800 \cdot \widetilde{B_0^{-1}} \cdot \tilde{f}_d^{-1} \cdot \tilde{f}_d^{-1} + 3.21900 \cdot \widetilde{N_{tot}^{-1}} \cdot 1 \cdot \widetilde{\Delta E_{F-s}} + (-4.45300) \cdot \tilde{f}_d^{-1} \cdot 1 \cdot \widetilde{W_d^{-1/2}}$	0.981

Table C.5: Results of the combinatorial search for the best fitting formula for nitrogen

Formula	ΔH function	correlation
$\Gamma_{(0,0,0)}$	$(-0.02700) + (-0.02700) + (-0.02700)$	0.000
$\Gamma_{(0,0,1)}$	$(-0.37000) + (-0.37000) + 0.65800 \cdot \widetilde{B}_0$	0.391
$\Gamma_{(0,0,2)}$	$(-0.79000) + (-0.79000) + 1.54400 \cdot \widetilde{W}_d^{1/2} \cdot \widetilde{R}_j$	0.486
$\Gamma_{(0,0,3)}$	$(-0.84200) + (-0.84200) + 2.14400 \cdot \widetilde{f}_d^{-1} \cdot \widetilde{R}_j \cdot \widetilde{\Delta E_{F-s}}$	0.538
$\Gamma_{(0,1,1)}$	$0.44500 + 1.54000 \cdot \widetilde{R}_j + (-2.06700) \cdot \widetilde{W}_d^{-1/2}$	0.498
$\Gamma_{(0,1,2)}$	$(-4.98600) + 1.00700 \cdot \widetilde{f}_d^{-1} + 4.01900 \cdot \widetilde{W}_d^{1/2} \cdot \widetilde{R}_j$	0.670
$\Gamma_{(0,1,3)}$	$(-0.19300) + (-1.87600) \cdot \widetilde{f}_d + 1.72900 \cdot \widetilde{R}_j \cdot \widetilde{R}_j \cdot \widetilde{\Delta E_{F-s}}$	0.751
$\Gamma_{(0,2,2)}$	$0.21300 + (-3.90300) \cdot \widetilde{W}_d^{-1/2} \cdot \widetilde{f}_d + 3.35000 \cdot \widetilde{R}_j \cdot \widetilde{f}_d$	0.728
$\Gamma_{(0,2,3)}$	$(-4.27200) + 2.76100 \cdot \widetilde{B}_0^{-1} \cdot \widetilde{R}_j + 2.31800 \cdot \widetilde{f}_d^{-1} \cdot \widetilde{\Delta E_{F-s}} \cdot \widetilde{B}_0$	0.767
$\Gamma_{(0,3,3)}$	$(-0.06900) + 2.34300 \cdot \widetilde{R}_j \cdot \widetilde{\Delta E_{F-s}} \cdot \widetilde{\Delta E_{F-s}} + (-3.13400) \cdot \widetilde{R}_j^{-1} \cdot \widetilde{B}_0 \cdot \widetilde{f}_d$	0.831
$\Gamma_{(1,1,1)}$	$3.40600 \cdot \widetilde{R}_j + (-4.23400) \cdot \widetilde{W}_d^{-1/2} + 0.76500 \cdot \widetilde{f}_d^{-1}$	0.644
$\Gamma_{(1,1,2)}$	$(-3.33700) \cdot \widetilde{f}_d + (-2.53000) \cdot \widetilde{B}_0 + 5.54700 \cdot \widetilde{R}_j \cdot \widetilde{\Delta E_{F-s}}$	0.854
$\Gamma_{(1,1,3)}$	$(-2.53000) \cdot \widetilde{B}_0 + (-3.33700) \cdot \widetilde{f}_d + 5.54700 \cdot 1 \cdot \widetilde{R}_j \cdot \widetilde{\Delta E_{F-s}}$	0.854
$\Gamma_{(1,2,2)}$	$(-3.29600) \cdot \widetilde{f}_d + (-2.90200) \cdot \widetilde{B}_0 \cdot \widetilde{N}_{tot} + 5.91600 \cdot \widetilde{R}_j \cdot \widetilde{\Delta E_{F-s}}$	0.860
$\Gamma_{(1,2,3)}$	$(-3.29600) \cdot \widetilde{f}_d + (-2.90200) \cdot \widetilde{B}_0 \cdot \widetilde{N}_{tot} + 5.91600 \cdot 1 \cdot \widetilde{R}_j \cdot \widetilde{\Delta E_{F-s}}$	0.860
$\Gamma_{(1,3,3)}$	$(-2.46700) \cdot \widetilde{B}_0 + 4.56600 \cdot \widetilde{R}_j \cdot \widetilde{\Delta E_{F-s}} \cdot \widetilde{\Delta E_{F-s}} + (-2.27900) \cdot \widetilde{\Delta E_{F-s}} \cdot \widetilde{\Delta E_{F-s}} \cdot \widetilde{f}_d$	0.876
$\Gamma_{(2,2,2)}$	$(-3.62200) \cdot \widetilde{W}_d^{1/2} \cdot \widetilde{f}_d + 5.66100 \cdot \widetilde{R}_j \cdot \widetilde{\Delta E_{F-s}} + (-2.52300) \cdot \widetilde{W}_d^{-1/2} \cdot \widetilde{B}_0$	0.866
$\Gamma_{(2,2,3)}$	$6.04800 \cdot \widetilde{R}_j \cdot \widetilde{\Delta E_{F-s}} + (-3.73300) \cdot \widetilde{W}_d^{1/2} \cdot \widetilde{f}_d + (-2.75700) \cdot \widetilde{W}_d^{-1/2} \cdot \widetilde{B}_0 \cdot \widetilde{N}_{tot}$	0.893
$\Gamma_{(2,3,3)}$	$(-3.20700) \cdot \widetilde{B}_0 \cdot \widetilde{N}_{tot} + 5.86100 \cdot \widetilde{R}_j \cdot \widetilde{\Delta E_{F-s}} \cdot \widetilde{B}_0 + (-2.91500) \cdot \widetilde{B}_0 \cdot \widetilde{B}_0 \cdot \widetilde{f}_d$	0.899
$\Gamma_{(3,3,3)}$	$(-3.20700) \cdot 1 \cdot \widetilde{B}_0 \cdot \widetilde{N}_{tot} + 5.86100 \cdot \widetilde{R}_j \cdot \widetilde{\Delta E_{F-s}} \cdot \widetilde{B}_0 + (-2.91500) \cdot \widetilde{B}_0 \cdot \widetilde{B}_0 \cdot \widetilde{f}_d$	0.899

Table C.6: Results of the combinatorial search for the best fitting formula for boron

Formula	ΔH function	correlation
$\Gamma_{(0,0,0)}$	$(-0.69400) + (-0.69400) + (-0.69400)$	0.000
$\Gamma_{(0,0,1)}$	$(-4.46300) + (-4.46300) + 6.84200 \cdot \widetilde{\Delta E_{F-s}}$	0.881
$\Gamma_{(0,0,2)}$	$(-5.23800) + (-5.23800) + 8.57600 \cdot \widetilde{W_d^{-1/2}} \cdot \widetilde{\Delta E_{F-s}}$	0.972
$\Gamma_{(0,0,3)}$	$(-5.23800) + (-5.23800) + 8.57600 \cdot \widetilde{W_d^{-1/2}} \cdot 1 \cdot \widetilde{\Delta E_{F-s}}$	0.972
$\Gamma_{(0,1,1)}$	$(-9.10900) + 3.89500 \cdot \widetilde{f_d} + 3.13100 \cdot \widetilde{B_0}$	0.970
$\Gamma_{(0,1,2)}$	$(-11.62000) + 0.41200 \cdot \widetilde{B_0^{-1}} + 9.32300 \cdot \widetilde{W_d^{-1/2}} \cdot \widetilde{\Delta E_{F-s}}$	0.974
$\Gamma_{(0,1,3)}$	$(-7.90600) + 3.32800 \cdot \widetilde{f_d} + 2.27000 \cdot \widetilde{W_d^{-1/2}} \cdot \widetilde{\Delta E_{F-s}} \cdot \widetilde{B_0}$	0.978
$\Gamma_{(0,2,2)}$	$(-14.97800) + 11.22300 \cdot \widetilde{W_d^{-1/2}} \cdot \widetilde{\Delta E_{F-s}} + 1.93700 \cdot \widetilde{\Delta E_{F-s}^{-1}} \cdot \widetilde{W_d^{1/2}}$	0.978
$\Gamma_{(0,2,3)}$	$(-15.19200) + 11.51200 \cdot \widetilde{W_d^{-1/2}} \cdot \widetilde{\Delta E_{F-s}} + 1.81600 \cdot \widetilde{\Delta E_{F-s}^{-1}} \cdot \widetilde{N_{tot}^{-1}} \cdot \widetilde{W_d^{1/2}}$	0.980
$\Gamma_{(0,3,3)}$	$(-6.20900) + (-6.13800) \cdot \widetilde{R_j} \cdot \widetilde{N_{tot}} \cdot \widetilde{N_{tot}} + 10.08900 \cdot \widetilde{W_d^{-1/2}} \cdot \widetilde{R_j} \cdot \widetilde{\Delta E_{F-s}}$	0.983
$\Gamma_{(1,1,1)}$	$3.89500 \cdot \widetilde{f_d} + (-9.10900) \cdot 1 + 3.13100 \cdot \widetilde{B_0}$	0.970
$\Gamma_{(1,1,2)}$	$3.19700 \cdot \widetilde{\Delta E_{F-s}} + (-7.97700) \cdot \widetilde{f_d^{-1}} + 1.82500 \cdot \widetilde{f_d^{-1}} \cdot \widetilde{f_d^{-1}}$	0.977
$\Gamma_{(1,1,3)}$	$(-14.25600) \cdot \widetilde{N_{tot}} + 1.14000 \cdot \widetilde{\Delta E_{F-s}^{-1}} + 11.06000 \cdot \widetilde{W_d^{-1/2}} \cdot \widetilde{\Delta E_{F-s}} \cdot \widetilde{N_{tot}}$	0.979
$\Gamma_{(1,2,2)}$	$(-17.81800) \cdot \widetilde{f_d^{-1}} + 5.11800 \cdot \widetilde{\Delta E_{F-s}^{-1}} \cdot \widetilde{f_d^{-1}} + 11.09900 \cdot \widetilde{f_d^{-1}} \cdot \widetilde{\Delta E_{F-s}}$	0.979
$\Gamma_{(1,2,3)}$	$1.80800 \cdot \widetilde{\Delta E_{F-s}} + (-7.55500) \cdot \widetilde{\Delta E_{F-s}^{-1}} \cdot \widetilde{f_d^{-1}} + 2.90400 \cdot \widetilde{\Delta E_{F-s}^{-1}} \cdot \widetilde{\Delta E_{F-s}^{-1}} \cdot \widetilde{f_d^{-1}}$	0.984
$\Gamma_{(1,3,3)}$	$(-15.84000) \cdot \widetilde{f_d^{-1}} + 10.08500 \cdot \widetilde{f_d^{-1}} \cdot \widetilde{\Delta E_{F-s}} \cdot \widetilde{N_{tot}} + 4.16400 \cdot \widetilde{\Delta E_{F-s}^{-1}} \cdot \widetilde{f_d^{-1}} \cdot \widetilde{N_{tot}}$	0.985
$\Gamma_{(2,2,2)}$	$11.09900 \cdot \widetilde{f_d^{-1}} \cdot \widetilde{\Delta E_{F-s}} + (-17.81800) \cdot \widetilde{f_d^{-1}} \cdot 1 + 5.11800 \cdot \widetilde{\Delta E_{F-s}^{-1}} \cdot \widetilde{f_d^{-1}}$	0.979
$\Gamma_{(2,2,3)}$	$1.18200 \cdot \widetilde{\Delta E_{F-s}} \cdot \widetilde{\Delta E_{F-s}} + (-6.63500) \cdot \widetilde{\Delta E_{F-s}^{-1}} \cdot \widetilde{f_d^{-1}} + 2.52400 \cdot \widetilde{\Delta E_{F-s}^{-1}} \cdot \widetilde{\Delta E_{F-s}^{-1}} \cdot \widetilde{f_d^{-1}}$	0.985
$\Gamma_{(2,3,3)}$	$(-18.40400) \cdot \widetilde{f_d^{-1}} \cdot 1 + 12.38900 \cdot \widetilde{W_d^{-1/2}} \cdot \widetilde{f_d^{-1}} \cdot \widetilde{\Delta E_{F-s}} + 4.71400 \cdot \widetilde{\Delta E_{F-s}^{-1}} \cdot \widetilde{f_d^{-1}} \cdot \widetilde{N_{tot}}$	0.991
$\Gamma_{(3,3,3)}$	$0.51100 \cdot \widetilde{B_0^{-1}} \cdot \widetilde{f_d^{-1}} \cdot \widetilde{f_d^{-1}} + 8.58000 \cdot \widetilde{f_d^{-1}} \cdot \widetilde{\Delta E_{F-s}} \cdot \widetilde{N_{tot}} + (-10.84200) \cdot \widetilde{f_d^{-1}} \cdot \widetilde{W_d^{1/2}} \cdot \widetilde{N_{tot}}$	0.992

Table C.7: Results of the combinatorial search for the best fitting formula for oxygen

Formula	ΔH function	correlation
$\Gamma_{(0,0,0)}$	$0.01000 + 0.01000 + 0.01000$	-0.000
$\Gamma_{(0,0,1)}$	$(-3.47800) + (-3.47800) + 6.98700 \cdot \widetilde{\Delta E_{F-s}}$	0.904
$\Gamma_{(0,0,2)}$	$(-2.20600) + (-2.20600) + 4.03800 \cdot \widetilde{B_0} \cdot \widetilde{f_d}$	0.964
$\Gamma_{(0,0,3)}$	$(-3.16700) + (-3.16700) + 6.28200 \cdot \widetilde{N_{tot}^{-1}} \cdot \widetilde{R_j} \cdot \widetilde{\Delta E_{F-s}}$	0.969
$\Gamma_{(0,1,1)}$	$(-7.03800) + 2.93300 \cdot \widetilde{f_d} + 4.13500 \cdot \widetilde{B_0}$	0.978
$\Gamma_{(0,1,2)}$	$(-7.03800) + 2.93300 \cdot \widetilde{f_d} + 4.13500 \cdot 1 \cdot \widetilde{B_0}$	0.978
$\Gamma_{(0,1,3)}$	$(-7.45700) + 4.41600 \cdot \widetilde{\Delta E_{F-s}} + 3.70800 \cdot \widetilde{\Delta E_{F-s}^{-1}} \cdot \widetilde{B_0} \cdot \widetilde{f_d}$	0.980
$\Gamma_{(0,2,2)}$	$(-7.51600) + 5.50000 \cdot \widetilde{W_d^{-1/2}} \cdot \widetilde{\Delta E_{F-s}} + 2.04100 \cdot \widetilde{R_j} \cdot \widetilde{B_0}$	0.979
$\Gamma_{(0,2,3)}$	$(-8.34500) + 1.75400 \cdot \widetilde{R_j} \cdot \widetilde{B_0} + 6.79600 \cdot \widetilde{W_d^{-1/2}} \cdot \widetilde{N_{tot}^{-1}} \cdot \widetilde{\Delta E_{F-s}}$	0.983
$\Gamma_{(0,3,3)}$	$(-8.61700) + 7.70400 \cdot \widetilde{W_d^{-1/2}} \cdot \widetilde{N_{tot}^{-1}} \cdot \widetilde{\Delta E_{F-s}} + 1.06700 \cdot \widetilde{R_j} \cdot \widetilde{B_0}$	0.984
$\Gamma_{(1,1,1)}$	$2.93300 \cdot \widetilde{f_d} + 4.13500 \cdot \widetilde{B_0} + (-7.03800) \cdot 1$	0.978
$\Gamma_{(1,1,2)}$	$3.82600 \cdot \widetilde{B_0} + (-5.68700) \cdot \widetilde{f_d^{-1}} + 1.25700 \cdot \widetilde{f_d^{-1}} \cdot \widetilde{f_d^{-1}}$	0.979
$\Gamma_{(1,1,3)}$	$4.41600 \cdot \widetilde{\Delta E_{F-s}} + (-7.45700) \cdot 1 + 3.70800 \cdot \widetilde{\Delta E_{F-s}^{-1}} \cdot \widetilde{B_0} \cdot \widetilde{f_d}$	0.980
$\Gamma_{(1,2,2)}$	$(-8.99900) \cdot \widetilde{N_{tot}} + 1.79400 \cdot \widetilde{R_j} \cdot \widetilde{B_0} + 7.29400 \cdot \widetilde{W_d^{-1/2}} \cdot \widetilde{\Delta E_{F-s}}$	0.983
$\Gamma_{(1,2,3)}$	$(-9.16200) \cdot \widetilde{N_{tot}} + 7.83400 \cdot \widetilde{W_d^{-1/2}} \cdot \widetilde{\Delta E_{F-s}} + 1.28600 \cdot \widetilde{R_j} \cdot \widetilde{B_0}$	0.987
$\Gamma_{(1,3,3)}$	$(-9.16200) \cdot \widetilde{N_{tot}} + 1.28600 \cdot \widetilde{R_j} \cdot \widetilde{B_0} + 7.83400 \cdot \widetilde{W_d^{-1/2}} \cdot 1 \cdot \widetilde{\Delta E_{F-s}}$	0.987
$\Gamma_{(2,2,2)}$	$7.29400 \cdot \widetilde{W_d^{-1/2}} \cdot \widetilde{\Delta E_{F-s}} + 1.79400 \cdot \widetilde{R_j} \cdot \widetilde{B_0} + (-8.99900) \cdot 1 \cdot \widetilde{N_{tot}}$	0.983
$\Gamma_{(2,2,3)}$	$(-9.16200) \cdot 1 \cdot \widetilde{N_{tot}} + 7.83400 \cdot \widetilde{W_d^{-1/2}} \cdot \widetilde{\Delta E_{F-s}} + 1.28600 \cdot \widetilde{R_j} \cdot \widetilde{B_0}$	0.987
$\Gamma_{(2,3,3)}$	$(-9.16200) \cdot 1 \cdot \widetilde{N_{tot}} + 7.77600 \cdot \widetilde{W_d^{-1/2}} \cdot \widetilde{W_d^{-1/2}} \cdot \widetilde{\Delta E_{F-s}} + 1.28600 \cdot \widetilde{R_j} \cdot \widetilde{B_0}$	0.987
$\Gamma_{(3,3,3)}$	$7.77600 \cdot \widetilde{W_d^{-1/2}} \cdot \widetilde{W_d^{-1/2}} \cdot \widetilde{\Delta E_{F-s}} + 1.28600 \cdot \widetilde{R_j} \cdot \widetilde{B_0} + (-9.16200) \cdot 1 \cdot 1 \cdot \widetilde{N_{tot}}$	0.987

Table C.8: Results of the combinatorial search for the best fitting formula for fluorine

Formula	ΔH function	correlation
$\Gamma_{(0,0,0)}$	$1.38700 + 1.38700 + 1.38700$	0.000
$\Gamma_{(0,0,1)}$	$0.67700 + 0.67700 + 2.80900 \cdot \widetilde{B}_0$	0.937
$\Gamma_{(0,0,2)}$	$0.68600 + 0.68600 + 2.76200 \cdot \widetilde{B}_0 \cdot \widetilde{N}_{tot}$	0.948
$\Gamma_{(0,0,3)}$	$1.18900 + 1.18900 + 1.45700 \cdot \widetilde{R}_j \cdot \widetilde{\Delta E_{F-s}} \cdot \widetilde{B}_0$	0.953
$\Gamma_{(0,1,1)}$	$1.02200 + 2.51200 \cdot \widetilde{B}_0 + 0.62800 \cdot \widetilde{f}_d$	0.956
$\Gamma_{(0,1,2)}$	$1.21600 + 1.37500 \cdot \widetilde{R}_j + 1.36300 \cdot \widetilde{\Delta E_{F-s}} \cdot \widetilde{B}_0$	0.972
$\Gamma_{(0,1,3)}$	$0.59500 + 1.98400 \cdot \widetilde{R}_j + 1.60700 \cdot \widetilde{R}_j^{-1} \cdot \widetilde{\Delta E_{F-s}} \cdot \widetilde{B}_0$	0.974
$\Gamma_{(0,2,2)}$	$0.90400 + 1.52000 \cdot \widetilde{R}_j \cdot \widetilde{B}_0 + 1.66800 \cdot \widetilde{N}_{tot}^{-1} \cdot \widetilde{\Delta E_{F-s}}$	0.974
$\Gamma_{(0,2,3)}$	$1.24600 + 4.69000 \cdot \widetilde{R}_j \cdot \widetilde{\Delta E_{F-s}} + (-1.87000) \cdot \widetilde{W}_d^{1/2} \cdot \widetilde{R}_j \cdot \widetilde{f}_d$	0.978
$\Gamma_{(0,3,3)}$	$0.58900 + 1.77900 \cdot \widetilde{W}_d^{-1/2} \cdot \widetilde{R}_j \cdot \widetilde{\Delta E_{F-s}} + 2.14500 \cdot \widetilde{f}_d^{-1} \cdot \widetilde{\Delta E_{F-s}} \cdot \widetilde{\Delta E_{F-s}}$	0.980
$\Gamma_{(1,1,1)}$	$(-1.06600) \cdot \widetilde{f}_d + 1.80200 \cdot \widetilde{R}_j + 3.46100 \cdot \widetilde{\Delta E_{F-s}}$	0.967
$\Gamma_{(1,1,2)}$	$(-1.47300) \cdot \widetilde{f}_d + 4.19900 \cdot \widetilde{\Delta E_{F-s}} + 1.42200 \cdot \widetilde{W}_d^{-1/2} \cdot \widetilde{R}_j$	0.980
$\Gamma_{(1,1,3)}$	$5.94300 \cdot \widetilde{\Delta E_{F-s}} + (-3.96300) \cdot \widetilde{f}_d + 1.80300 \cdot \widetilde{W}_d^{-1/2} \cdot \widetilde{R}_j \cdot \widetilde{f}_d$	0.986
$\Gamma_{(1,2,2)}$	$5.16900 \cdot \widetilde{\Delta E_{F-s}} + (-2.15700) \cdot \widetilde{W}_d^{1/2} \cdot \widetilde{f}_d + 0.99300 \cdot \widetilde{R}_j \cdot \widetilde{R}_j$	0.985
$\Gamma_{(1,2,3)}$	$6.05200 \cdot \widetilde{\Delta E_{F-s}} + 1.83900 \cdot \widetilde{R}_j \cdot \widetilde{f}_d + (-3.96400) \cdot \widetilde{W}_d^{1/2} \cdot \widetilde{f}_d \cdot \widetilde{N}_{tot}$	0.992
$\Gamma_{(1,3,3)}$	$5.74600 \cdot \widetilde{\Delta E_{F-s}} + 2.08300 \cdot \widetilde{N}_{tot}^{-1} \cdot \widetilde{R}_j \cdot \widetilde{f}_d + (-3.82400) \cdot \widetilde{W}_d^{1/2} \cdot \widetilde{f}_d \cdot \widetilde{N}_{tot}$	0.992
$\Gamma_{(2,2,2)}$	$5.77500 \cdot \widetilde{N}_{tot}^{-1} \cdot \widetilde{\Delta E_{F-s}} + 1.82900 \cdot \widetilde{R}_j \cdot \widetilde{f}_d + (-3.66500) \cdot \widetilde{W}_d^{1/2} \cdot \widetilde{f}_d$	0.987
$\Gamma_{(2,2,3)}$	$6.05200 \cdot 1 \cdot \widetilde{\Delta E_{F-s}} + 1.83900 \cdot \widetilde{R}_j \cdot \widetilde{f}_d + (-3.96400) \cdot \widetilde{W}_d^{1/2} \cdot \widetilde{f}_d \cdot \widetilde{N}_{tot}$	0.992
$\Gamma_{(2,3,3)}$	$5.59300 \cdot \widetilde{W}_d^{-1/2} \cdot \widetilde{\Delta E_{F-s}} + (-3.44900) \cdot \widetilde{f}_d \cdot \widetilde{N}_{tot} \cdot \widetilde{W}_d^{-1/2} \cdot \widetilde{R}_j \cdot \widetilde{f}_d$	0.994
$\Gamma_{(3,3,3)}$	$5.59300 \cdot \widetilde{W}_d^{-1/2} \cdot 1 \cdot \widetilde{\Delta E_{F-s}} + (-3.44900) \cdot \widetilde{f}_d \cdot \widetilde{N}_{tot} \cdot \widetilde{W}_d^{-1/2} \cdot \widetilde{R}_j \cdot \widetilde{f}_d$	0.994

Table C.9: Results of the combinatorial search for the best fitting formula for helium

Formula	ΔH function	correlation
$\Gamma_{(0,0,0)}$	$2.47200 + 2.47200 + 2.47200$	0.000
$\Gamma_{(0,0,1)}$	$0.98200 + 0.98200 + 5.45200 \cdot \widetilde{B}_0$	0.951
$\Gamma_{(0,0,2)}$	$0.98200 + 0.98200 + 5.45200 \cdot 1 \cdot \widetilde{B}_0$	0.951
$\Gamma_{(0,0,3)}$	$1.22500 + 1.22500 + 4.80000 \cdot \widetilde{N}_{tot}^{-1} \cdot \widetilde{R}_j \cdot \widetilde{B}_0$	0.963
$\Gamma_{(0,1,1)}$	$0.92900 + 5.04300 \cdot \widetilde{B}_0 + 1.44300 \cdot \widetilde{R}_j$	0.964
$\Gamma_{(0,1,2)}$	$10.01100 + (-8.01600) \cdot \widetilde{N}_{tot} + 5.11400 \cdot \widetilde{R}_j \cdot \widetilde{B}_0$	0.966
$\Gamma_{(0,1,3)}$	$0.76300 + 1.70600 \cdot \widetilde{W}_d^{1/2} + 4.78100 \cdot \widetilde{N}_{tot}^{-1} \cdot \widetilde{R}_j \cdot \widetilde{B}_0$	0.969
$\Gamma_{(0,2,2)}$	$0.76400 + 4.42800 \cdot \widetilde{W}_d^{1/2} \cdot \widetilde{B}_0 + 2.06900 \cdot \widetilde{W}_d^{-1/2} \cdot \widetilde{R}_j$	0.969
$\Gamma_{(0,2,3)}$	$3.06300 + 6.58000 \cdot \widetilde{R}_j \cdot \widetilde{\Delta E}_{F-s} + (-3.22500) \cdot \widetilde{B}_0^{-1} \cdot \widetilde{\Delta E}_{F-s} \cdot \widetilde{f}_d$	0.977
$\Gamma_{(0,3,3)}$	$2.55900 + 6.77300 \cdot \widetilde{N}_{tot}^{-1} \cdot \widetilde{R}_j \cdot \widetilde{\Delta E}_{F-s} + (-2.59400) \cdot \widetilde{B}_0^{-1} \cdot \widetilde{\Delta E}_{F-s} \cdot \widetilde{f}_d$	0.985
$\Gamma_{(1,1,1)}$	$5.24800 \cdot \widetilde{B}_0 + 1.84300 \cdot \widetilde{R}_j + 0.32900 \cdot \widetilde{f}_d^{-1}$	0.965
$\Gamma_{(1,1,2)}$	$4.83500 \cdot \widetilde{B}_0 + 1.91500 \cdot \widetilde{R}_j + 0.84500 \cdot \widetilde{f}_d^{-1} \cdot \widetilde{\Delta E}_{F-s}$	0.968
$\Gamma_{(1,1,3)}$	$3.43900 \cdot \widetilde{R}_j + 6.19900 \cdot \widetilde{\Delta E}_{F-s} + (-2.80500) \cdot \widetilde{B}_0^{-1} \cdot \widetilde{\Delta E}_{F-s} \cdot \widetilde{f}_d$	0.976
$\Gamma_{(1,2,2)}$	$7.31700 \cdot \widetilde{f}_d + (-4.74400) \cdot \widetilde{f}_d \cdot \widetilde{f}_d + 5.69400 \cdot \widetilde{R}_j \cdot \widetilde{\Delta E}_{F-s}$	0.979
$\Gamma_{(1,2,3)}$	$2.93900 \cdot \widetilde{R}_j + 11.86000 \cdot \widetilde{\Delta E}_{F-s} \cdot \widetilde{f}_d + (-6.14700) \cdot \widetilde{\Delta E}_{F-s} \cdot \widetilde{f}_d \cdot \widetilde{f}_d$	0.984
$\Gamma_{(1,3,3)}$	$5.73100 \cdot \widetilde{\Delta E}_{F-s} + (-1.66400) \cdot \widetilde{f}_d \cdot \widetilde{f}_d + 4.24800 \cdot \widetilde{N}_{tot}^{-1} \cdot \widetilde{R}_j \cdot \widetilde{f}_d$	0.986
$\Gamma_{(2,2,2)}$	$6.41100 \cdot \widetilde{N}_{tot}^{-1} \cdot \widetilde{f}_d + (-3.86000) \cdot \widetilde{f}_d \cdot \widetilde{f}_d + 5.71700 \cdot \widetilde{R}_j \cdot \widetilde{\Delta E}_{F-s}$	0.980
$\Gamma_{(2,2,3)}$	$(-4.50500) \cdot \widetilde{f}_d \cdot \widetilde{f}_d + 6.26800 \cdot \widetilde{W}_d^{-1/2} \cdot \widetilde{f}_d + 6.23400 \cdot \widetilde{W}_d^{1/2} \cdot \widetilde{R}_j \cdot \widetilde{\Delta E}_{F-s}$	0.987
$\Gamma_{(2,3,3)}$	$11.91400 \cdot 1 \cdot \widetilde{f}_d + 3.13000 \cdot \widetilde{R}_j \cdot \widetilde{\Delta E}_{F-s} \cdot \widetilde{\Delta E}_{F-s} + (-6.90400) \cdot \widetilde{W}_d^{1/2} \cdot \widetilde{f}_d \cdot \widetilde{f}_d$	0.990
$\Gamma_{(3,3,3)}$	$11.46200 \cdot \widetilde{W}_d^{-1/2} \cdot \widetilde{N}_{tot}^{-1} \cdot \widetilde{f}_d + 3.35300 \cdot \widetilde{R}_j \cdot \widetilde{\Delta E}_{F-s} \cdot \widetilde{\Delta E}_{F-s} + (-6.71100) \cdot \widetilde{N}_{tot}^{-1} \cdot \widetilde{f}_d \cdot \widetilde{f}_d$	0.991

Table C.10: Results of the combinatorial search for the best fitting formula for neon

Formula	ΔH function	correlation
$\Gamma_{(0,0,0)}$	$0.00900 + 0.00900 + 0.00900$	0.000
$\Gamma_{(0,0,1)}$	$(-0.60800) + (-0.60800) + 1.24500 \cdot \widetilde{\Delta E_{F-s}^{-1}}$	0.854
$\Gamma_{(0,0,2)}$	$(-0.49100) + (-0.49100) + 0.97100 \cdot \widetilde{R_j} \cdot \widetilde{\Delta E_{F-s}^{-1}}$	0.916
$\Gamma_{(0,0,3)}$	$(-0.49100) + (-0.49100) + 0.97100 \cdot 1 \cdot \widetilde{R_j} \cdot \widetilde{\Delta E_{F-s}^{-1}}$	0.916
$\Gamma_{(0,1,1)}$	$(-1.23700) + 0.12200 \cdot \widetilde{f_d} + 1.14300 \cdot \widetilde{\Delta E_{F-s}^{-1}}$	0.911
$\Gamma_{(0,1,2)}$	$(-0.77900) + (-0.08800) \cdot \widetilde{B_0^{-1}} + 0.86100 \cdot \widetilde{R_j} \cdot \widetilde{\Delta E_{F-s}^{-1}}$	0.939
$\Gamma_{(0,1,3)}$	$(-1.01600) + (-0.01900) \cdot \widetilde{B_0^{-1}} + 1.05000 \cdot \widetilde{N_{tot}^{-1}} \cdot \widetilde{R_j} \cdot \widetilde{\Delta E_{F-s}^{-1}}$	0.940
$\Gamma_{(0,2,2)}$	$(-0.81100) + 0.87400 \cdot \widetilde{R_j} \cdot \widetilde{\Delta E_{F-s}^{-1}} + (-0.06800) \cdot \widetilde{B_0^{-1}} \cdot \widetilde{N_{tot}^{-1}}$	0.939
$\Gamma_{(0,2,3)}$	$(-1.00200) + (-0.02700) \cdot \widetilde{B_0^{-1}} \cdot \widetilde{N_{tot}^{-1}} + 1.04400 \cdot \widetilde{N_{tot}^{-1}} \cdot \widetilde{R_j} \cdot \widetilde{\Delta E_{F-s}^{-1}}$	0.942
$\Gamma_{(0,3,3)}$	$(-0.98700) + 1.03800 \cdot \widetilde{N_{tot}^{-1}} \cdot \widetilde{R_j} \cdot \widetilde{\Delta E_{F-s}^{-1}} + (-0.03800) \cdot \widetilde{B_0^{-1}} \cdot \widetilde{N_{tot}^{-1}} \cdot \widetilde{N_{tot}^{-1}}$	0.944
$\Gamma_{(1,1,1)}$	$(-0.80100) \cdot \widetilde{W_d^{1/2}} + 1.01000 \cdot \widetilde{\Delta E_{F-s}^{-1}} + (-0.16900) \cdot \widetilde{B_0^{-1}}$	0.915
$\Gamma_{(1,1,2)}$	$(-0.77900) \cdot 1 + (-0.08800) \cdot \widetilde{B_0^{-1}} + 0.86100 \cdot \widetilde{R_j} \cdot \widetilde{\Delta E_{F-s}^{-1}}$	0.939
$\Gamma_{(1,1,3)}$	$(-1.01600) \cdot 1 + (-0.01900) \cdot \widetilde{B_0^{-1}} + 1.05000 \cdot \widetilde{N_{tot}^{-1}} \cdot \widetilde{R_j} \cdot \widetilde{\Delta E_{F-s}^{-1}}$	0.940
$\Gamma_{(1,2,2)}$	$(-0.81100) \cdot 1 + 0.87400 \cdot \widetilde{R_j} \cdot \widetilde{\Delta E_{F-s}^{-1}} + (-0.06800) \cdot \widetilde{B_0^{-1}} \cdot \widetilde{N_{tot}^{-1}}$	0.939
$\Gamma_{(1,2,3)}$	$(-1.00200) \cdot 1 + (-0.02700) \cdot \widetilde{B_0^{-1}} \cdot \widetilde{N_{tot}^{-1}} + 1.04400 \cdot \widetilde{N_{tot}^{-1}} \cdot \widetilde{R_j} \cdot \widetilde{\Delta E_{F-s}^{-1}}$	0.942
$\Gamma_{(1,3,3)}$	$0.21000 \cdot \widetilde{B_0^{-1}} + 0.64700 \cdot \widetilde{N_{tot}^{-1}} \cdot \widetilde{R_j} \cdot \widetilde{\Delta E_{F-s}^{-1}} + (-0.86700) \cdot \widetilde{\Delta E_{F-s}^{-1}} \cdot \widetilde{W_d^{1/2}} \cdot \widetilde{N_{tot}^{-1}}$	0.945
$\Gamma_{(2,2,2)}$	$0.87400 \cdot \widetilde{R_j} \cdot \widetilde{\Delta E_{F-s}^{-1}} + (-1.13600) \cdot \widetilde{f_d^{-1}} \cdot \widetilde{f_d} + (-0.06800) \cdot \widetilde{B_0^{-1}} \cdot \widetilde{N_{tot}^{-1}}$	0.939
$\Gamma_{(2,2,3)}$	$0.02900 \cdot \widetilde{B_0^{-1}} \cdot \widetilde{B_0^{-1}} + 0.59500 \cdot \widetilde{R_j} \cdot \widetilde{\Delta E_{F-s}^{-1}} + (-0.66400) \cdot \widetilde{\Delta E_{F-s}^{-1}} \cdot \widetilde{W_d^{1/2}} \cdot \widetilde{N_{tot}^{-1}}$	0.947
$\Gamma_{(2,3,3)}$	$(-0.70800) \cdot \widetilde{\Delta E_{F-s}^{-1}} \cdot \widetilde{W_d^{1/2}} + 0.62600 \cdot \widetilde{N_{tot}^{-1}} \cdot \widetilde{R_j} \cdot \widetilde{\Delta E_{F-s}^{-1}} + 0.07400 \cdot \widetilde{W_d^{-1/2}} \cdot \widetilde{\Delta E_{F-s}^{-1}} \cdot \widetilde{B_0^{-1}}$	0.948
$\Gamma_{(3,3,3)}$	$(-1.06400) \cdot \widetilde{R_j^{-1}} \cdot \widetilde{f_d^{-1}} \cdot \widetilde{B_0^{-1}} + 1.11000 \cdot \widetilde{f_d^{-1}} \cdot \widetilde{\Delta E_{F-s}^{-1}} \cdot \widetilde{\Delta E_{F-s}^{-1}} + (-0.00239) \cdot \widetilde{\Delta E_{F-s}^{-1}} \cdot \widetilde{\Delta E_{F-s}^{-1}} \cdot \widetilde{B_0^{-1}}$	0.950

Table C.11: Results of the combinatorial search for group I elements containing hydrogen, carbon, nitrogen, and oxygen. The coefficients shown here corresponds to the interstitial element hydrogen, whereas the correlation value is for the average of all four interstitial elements.

Formula	ΔH function	correlation
$\Gamma_{(0,0,0)}$	$0.01000 + 0.01000 + 0.01000$	0.000
$\Gamma_{(0,0,1)}$	$(-2.74600) + (-2.74600) + 5.52300 \cdot \widetilde{B}_0$	0.922
$\Gamma_{(0,0,2)}$	$(-2.80500) + (-2.80500) + 5.58300 \cdot \widetilde{B}_0 \cdot \widetilde{N}_{tot}$	0.937
$\Gamma_{(0,0,3)}$	$(-2.79400) + (-2.79400) + 5.47700 \cdot \widetilde{B}_0 \cdot \widetilde{N}_{tot} \cdot \widetilde{N}_{tot}$	0.938
$\Gamma_{(0,1,1)}$	$(-7.03800) + 4.13500 \cdot \widetilde{B}_0 + 2.93300 \cdot \widetilde{f}_d$	0.962
$\Gamma_{(0,1,2)}$	$(-6.25000) + 1.59900 \cdot \widetilde{B}_0 + 4.50000 \cdot \widetilde{R}_j \cdot \widetilde{\Delta E_{F-s}}$	0.966
$\Gamma_{(0,1,3)}$	$(-6.38400) + 0.64700 \cdot \widetilde{B}_0 + 5.69400 \cdot \widetilde{N}_{tot}^{-1} \cdot \widetilde{R}_j \cdot \widetilde{\Delta E_{F-s}}$	0.967
$\Gamma_{(0,2,2)}$	$(-5.39900) + 5.87300 \cdot \widetilde{R}_j \cdot \widetilde{\Delta E_{F-s}} + (-0.86800) \cdot \widetilde{B}_0^{-1} \cdot \widetilde{f}_d$	0.970
$\Gamma_{(0,2,3)}$	$(-5.95800) + 6.14700 \cdot \widetilde{R}_j \cdot \widetilde{\Delta E_{F-s}} + (-0.45200) \cdot \widetilde{B}_0^{-1} \cdot \widetilde{f}_d \cdot \widetilde{f}_d$	0.972
$\Gamma_{(0,3,3)}$	$(-6.29500) + 6.39400 \cdot \widetilde{N}_{tot}^{-1} \cdot \widetilde{R}_j \cdot \widetilde{\Delta E_{F-s}} + (-0.19600) \cdot \widetilde{B}_0^{-1} \cdot \widetilde{\Delta E_{F-s}} \cdot \widetilde{f}_d$	0.978
$\Gamma_{(1,1,1)}$	$(-7.03800) \cdot 1 + 4.13500 \cdot \widetilde{B}_0 + 2.93300 \cdot \widetilde{f}_d$	0.962
$\Gamma_{(1,1,2)}$	$(-6.11300) \cdot \widetilde{N}_{tot}^{-1} + 3.72300 \cdot \widetilde{\Delta E_{F-s}} + 2.28800 \cdot \widetilde{R}_j \cdot \widetilde{B}_0$	0.969
$\Gamma_{(1,1,3)}$	$3.73600 \cdot \widetilde{\Delta E_{F-s}} + (-6.19600) \cdot \widetilde{N}_{tot}^{-1} + 2.40200 \cdot \widetilde{N}_{tot}^{-1} \cdot \widetilde{R}_j \cdot \widetilde{B}_0$	0.971
$\Gamma_{(1,2,2)}$	$(-5.39900) \cdot 1 + 5.87300 \cdot \widetilde{R}_j \cdot \widetilde{\Delta E_{F-s}} + (-0.86800) \cdot \widetilde{B}_0^{-1} \cdot \widetilde{f}_d$	0.970
$\Gamma_{(1,2,3)}$	$(-5.95800) \cdot 1 + 6.14700 \cdot \widetilde{R}_j \cdot \widetilde{\Delta E_{F-s}} + (-0.45200) \cdot \widetilde{B}_0^{-1} \cdot \widetilde{f}_d \cdot \widetilde{f}_d$	0.972
$\Gamma_{(1,3,3)}$	$(-6.29500) \cdot 1 + 6.39400 \cdot \widetilde{N}_{tot}^{-1} \cdot \widetilde{R}_j \cdot \widetilde{\Delta E_{F-s}} + (-0.19600) \cdot \widetilde{B}_0^{-1} \cdot \widetilde{\Delta E_{F-s}} \cdot \widetilde{f}_d$	0.978
$\Gamma_{(2,2,2)}$	$(-0.86800) \cdot \widetilde{B}_0^{-1} \cdot \widetilde{f}_d + (-6.17000) \cdot \widetilde{R}_j^{-1} \cdot \widetilde{R}_j + 5.87300 \cdot \widetilde{R}_j \cdot \widetilde{\Delta E_{F-s}}$	0.970
$\Gamma_{(2,2,3)}$	$0.21500 \cdot \widetilde{R}_j \cdot \widetilde{R}_j + 0.62700 \cdot \widetilde{N}_{tot}^{-1} \cdot \widetilde{\Delta E_{F-s}} + (-0.02900) \cdot \widetilde{B}_0^{-1} \cdot \widetilde{\Delta E_{F-s}} \cdot \widetilde{f}_d$	0.975
$\Gamma_{(2,3,3)}$	$0.18800 \cdot \widetilde{R}_j \cdot \widetilde{R}_j + (-0.01000) \cdot \widetilde{B}_0^{-1} \cdot \widetilde{\Delta E_{F-s}} \cdot \widetilde{f}_d + 0.64100 \cdot \widetilde{W}_d^{-1/2} \cdot \widetilde{N}_{tot}^{-1} \cdot \widetilde{\Delta E_{F-s}}$	0.979
$\Gamma_{(3,3,3)}$	$0.69900 \cdot \widetilde{W}_d^{-1/2} \cdot \widetilde{N}_{tot}^{-1} \cdot \widetilde{\Delta E_{F-s}} + 0.02200 \cdot \widetilde{B}_0^{-1} \cdot \widetilde{\Delta E_{F-s}} \cdot \widetilde{f}_d + 0.10800 \cdot \widetilde{W}_d^{-1/2} \cdot \widetilde{R}_j \cdot \widetilde{R}_j$	0.981

Table C.12: Results of the combinatorial search for group II elements containing the elements fluorine, helium, and neon. The coefficients shown here corresponds to the interstitial element fluorine, whereas the correlation value is for the average of all three interstitial elements.

Bibliography

- [1] J. Warren and R. Boisvert, *Building the Materials Innovation Infrastructure: Data and Standards. A Materials Genome Initiative Workshop* (National Institute of Standards and Technology, Gaithersburg, MD, 2012).
- [2] W. F. Maier, K. Stowe, and S. Sieg, *Angew Chem Int Ed Engl* **46**, 6016 (2007), ISSN 1433-7851 (Print) 1433-7851 (Linking).
- [3] W. Dickson and A. Dickson, *The Life and Inventions of Thomas Alva Edison* (T. Y. Crowell, 1894), URL <https://books.google.de/books?id=qUY0AAAAAYAAJ>.
- [4] T. Edison, *Electric lamp* (1880), uS Patent 223,898, URL <http://www.google.com/patents/US223898>.
- [5] R. Hoogenboom, M. A. R. Meier, and U. S. Schubert, *Macromolecular Rapid Communications* **24**, 16 (2003), ISSN 1022-1336.
- [6] (2015-12-07), URL http://www.researchgate.net/figure/228021377_fig12_Figure-1-Ciamician-tested-hundreds-of-samples-in-parallel-on-the-roof-of-his-laboratory.
- [7] O. Z. Fisher, A. Khademhosseini, R. Langer, and N. A. Peppas, *Accounts of chemical research* **43**, 419 (2009).
- [8] A. Ludwig, J. Cao, A. Savan, and M. Ehmann, *Journal of Alloys and Compounds* **446**, 516 (2007), ISSN 0925-8388.
- [9] J. Cui, Y. S. Chu, O. O. Famodu, Y. Furuya, J. Hattrick-Simpers, R. D. James, A. Ludwig, S. Thienhaus, M. Wuttig, Z. Zhang, et al., *Nat Mater* **5**, 286 (2006).
- [10] L. Kaufman and H. Bernstein (1970).
- [11] L. Kaufman and J. Ågren, *Scripta Materialia* **70**, 3 (2014), ISSN 1359-6462.
- [12] M. Born and R. Oppenheimer, *Annalen der Physik* **389**, 457 (1927), ISSN 1521-3889.
- [13] P. Hohenberg and W. Kohn, *Physical Review* **136**, B864 (1964).
- [14] R. Nazarov, R. Hood, and M. Morales, in *APS Meeting Abstracts* (2015), vol. 1, p. 10011.

- [15] K. Lejaeghere, V. Van Speybroeck, G. Van Oost, and S. Cottenier, *Critical Reviews in Solid State and Materials Sciences* **39**, 1 (2014), ISSN 1040-8436.
- [16] M. S. Daw and M. I. Baskes, *Physical Review Letters* **50**, 1285 (1983).
- [17] M. S. Daw and M. I. Baskes, *Physical Review B* **29**, 6443 (1984).
- [18] M. S. Daw, S. M. Foiles, and M. I. Baskes, *Materials Science Reports* **9**, 251 (1993).
- [19] M. Baskes, *Physical review letters* **59**, 2666 (1987).
- [20] M. Baskes, J. Nelson, and A. Wright, *Physical Review B* **40**, 6085 (1989).
- [21] J. Hafner, C. Wolverton, and G. Ceder, *MRS bulletin* **31**, 659 (2006).
- [22] S. Curtarolo, G. L. Hart, M. B. Nardelli, N. Mingo, S. Sanvito, and O. Levy, *Nat Mater* **12**, 191 (2013), ISSN 1476-1122 (Print) 1476-1122 (Linking).
- [23] S. Srinivasan, S. R. Broderick, R. Zhang, A. Mishra, S. B. Sinnott, S. K. Saxena, J. M. LeBeau, and K. Rajan, *Scientific reports* **5** (2015).
- [24] A. Miedema, P. De Chatel, and F. De Boer, *Physica B+ C* **100**, 1 (1980).
- [25] A. Miedema, A. Niessen, F. De Boer, R. Boom, and W. Matten, Report, Philips Research Laboratories, Eindhoven, The Netherlands. FR de Boer, R. Boom, WCM Mattens, AR Miedema, AK Niessen: Cohesion in Metals. Transition Metals Alloys, North-Holland Publishing Co., Amsterdam (1989).
- [26] W. B. Pearson (1972).
- [27] B. Seiser, R. Drautz, and D. Pettifor, *Acta Materialia* **59**, 749 (2011).
- [28] S. Sandlöbes, M. Friák, S. Zaefferer, A. Dick, S. Yi, D. Letzig, Z. Pei, L.-F. Zhu, J. Neugebauer, and D. Raabe, *Acta Materialia* **60**, 3011 (2012).
- [29] A. Abbasi, A. Dick, T. Hickel, and J. Neugebauer, *Acta Materialia* **59**, 3041 (2011).
- [30] A. G. Van der Geest and A. N. Kolmogorov, *Calphad-Computer Coupling of Phase Diagrams and Thermochemistry* **46**, 184 (2014).
- [31] K. F. Garrity, J. W. Bennett, K. M. Rabe, and D. Vanderbilt, *Computational Materials Science* **81**, 446 (2014).
- [32] R. Armiento, B. Kozinsky, M. Fornari, and G. Ceder, *Physical Review B* **84** (2011).
- [33] A. Jain, G. Hautier, C. J. Moore, S. P. Ong, C. C. Fischer, T. Mueller, K. A. Persson, and G. Ceder, *Computational Materials Science* **50**, 2295 (2011), ISSN 0927-0256.
- [34] G. Hautier, A. Jain, S. P. Ong, B. Kang, C. Moore, R. Doe, and G. Ceder, *Chemistry of Materials* **23**, 3495 (2011), ISSN 0897-4756.
- [35] J. E. Saal, S. Kirklin, M. Aykol, B. Meredig, and C. Wolverton, *Jom* **65**, 1501 (2013).

- [36] N. Mott, in *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* (The Royal Society, 1936), vol. 153, pp. 699–717.
- [37] J. Zaanen, G. Sawatzky, and J. Allen, *Physical Review Letters* **55**, 418 (1985).
- [38] F. Koermann, A. Dick, T. Hickel, and J. Neugebauer, *Physical Review B* **83** (2011), ISSN 1098-0121.
- [39] J. Wills, O. Eriksson, P. Söderlind, and A. Boring, *Physical review letters* **68**, 2802 (1992).
- [40] D. E. Jiang and E. A. Carter, *Physical Review B* **70** (2004).
- [41] T. Hickel, S. Sandlöbes, R. K. Marceau, A. Dick, I. Bleskov, J. Neugebauer, and D. Raabe, *Acta materialia* **75**, 147 (2014).
- [42] T. Gladman, *Materials science and technology* **15**, 30 (1999).
- [43] D. Jack and K. Jack, *Materials Science and Engineering* **11**, 1 (1973).
- [44] I. Robertson, *Engineering Fracture Mechanics* **68**, 671 (2001).
- [45] D. Caillard, *Acta Materialia* **59**, 4974 (2011).
- [46] C. D. Beachem, *Metallurgical Transactions* **3**, 437 (1972).
- [47] H. K. Birnbaum and P. Sofronis, *Materials Science and Engineering a-Structural Materials Properties Microstructure and Processing* **176**, 191 (1994).
- [48] C. G. McKamey, J. H. Devan, P. F. Tortorelli, and V. K. Sikka, *Journal of Materials Research* **6**, 1779 (1991).
- [49] V. G. Gavriljuk, *Isij International* **36**, 738 (1996).
- [50] Y. Tomota, Y. Xia, and K. Inoue, *Acta Materialia* **46**, 1577 (1998).
- [51] C. H. Olk, *Measurement Science and Technology* **16**, 14 (2005).
- [52] R. Gremaud, M. Slaman, H. Schreuders, B. Dam, and R. Griessen, *Applied Physics Letters* **91** (2007), ISSN 0003-6951.
- [53] R. Gremaud, A. Borgschulte, C. Chacon, J. L. M. Van Mechelen, H. Schreuders, A. Zuttel, B. Hjorvarsson, B. Dam, and R. Griessen, *Applied Physics a-Materials Science & Processing* **84**, 77 (2006), ISSN 0947-8396.
- [54] A. Borgschulte, R. J. Westerwaal, J. H. Rector, B. Dam, and R. Griessen, *Applied Physics Letters* **85**, 4884 (2004).
- [55] H. Wipf, *Physica Scripta* **T94**, 43-51, 2001 (2001).
- [56] J. J. Hanak, *Journal of Materials Science* **5**, 964 (1970), ISSN 0022-2461.
- [57] G. Ceder, *Mrs Bulletin* **35**, 693 (2010), ISSN 0883-7694.

- [58] L. Vegard, *Zeitschrift Fur Physik* **5**, 17 (1921).
- [59] B. Jandeleit, D. J. Schaefer, T. S. Powers, H. W. Turner, and W. H. Weinberg, *Angewandte Chemie-International Edition* **38**, 2495 (1999), ISSN 1433-7851.
- [60] X. D. Xiang, X. D. Sun, G. Briceno, Y. L. Lou, K. A. Wang, H. Y. Chang, W. G. Wallace-freedman, S. W. Chen, and P. G. Schultz, *Science* **268**, 1738 (1995), ISSN 0036-8075.
- [61] H. Chang, C. Gao, I. Takeuchi, Y. Yoo, J. Wang, P. G. Schultz, X. D. Xiang, R. P. Sharma, M. Downes, and T. Venkatesan, *Applied Physics Letters* **72**, 2185 (1998), ISSN 0003-6951.
- [62] G. Briceno, H. Y. Chang, X. D. Sun, P. G. Schultz, and X. D. Xiang, *Science* **270**, 273 (1995).
- [63] J. C. Zhao, *Advanced Engineering Materials* **3**, 143 (2001).
- [64] C. H. Olk, *Measurement Science & Technology* **16**, 14 (2005).
- [65] I. Takeuchi, O. O. Famodu, J. C. Read, M. A. Aronova, K. S. Chang, C. Craciunescu, S. E. Loffland, M. Wuttig, F. C. Wellstood, L. Knauss, et al., *Nature Materials* **2**, 180 (2003).
- [66] A. Holzwarth, P. W. Schmidt, and W. E. Maier, *Angewandte Chemie-International Edition* **37**, 2644 (1998).
- [67] A. C. Cooper, L. H. McAlexander, D. H. Lee, M. T. Torres, and R. H. Crabtree, *Journal of the American Chemical Society* **120**, 9971 (1998).
- [68] J. P. Lemmon, R. J. Wroczynski, D. W. Whisenhunt, and W. P. Flanagan, *Abstracts of Papers of the American Chemical Society* **222**, U307 (2001).
- [69] P. Strasser, Q. Fan, M. Devenney, W. H. Weinberg, P. Liu, and J. K. Nørskov, *Journal of Physical Chemistry B* **107**, 11013 (2003).
- [70] R. Z. Jiang, C. Rong, and D. Chu, *Journal of Combinatorial Chemistry* **7**, 272 (2005).
- [71] J. P. Lemmon, V. Manivannan, T. Jordan, L. Hassib, T. Siclovan, M. Othon, and M. Pilliod, *High throughput screening of materials for solid oxide fuel cells* (Dunno, 2004), vol. 804 of *Materials Research Society Symposium Proceedings*, pp. 27–32.
- [72] H. Hansel, H. Zettl, G. Krausch, C. Schmitz, R. Kisselev, M. Thelakkat, and H. W. Schmidt, *Applied Physics Letters* **81**, 2106 (2002).
- [73] B. Chisholm, R. Potyrailo, J. Cawse, R. Shaffer, M. Brennan, C. Molaison, D. Whisenhunt, B. Flanagan, D. Olson, J. Akhave, et al., *Progress in Organic Coatings* **45**, 313 (2002).
- [74] D. D. Landis, J. S. Hummelshøj, S. Nestorov, J. Greeley, M. Dulak, T. Bligaard, J. K. Nørskov, and K. W. Jacobsen, *Computing in Science & Engineering* **14**, 51 (2012).
- [75] S. Curtarolo, W. Setyawan, S. Wang, J. Xue, K. Yang, R. H. Taylor, L. J. Nelson, G. L. W. Hart, S. Sanvito, M. Buongiorno Nardelli, et al., *Computational Materials Science* **58**, 227 (2012).

- [76] G. H. Johannesson, T. Bligaard, A. V. Ruban, H. L. Skriver, K. W. Jacobsen, and J. K. Nørskov, *Physical Review Letters* **88** (2002), ISSN 0031-9007.
- [77] D. P. Stucke and V. H. Crespi, *Nano Letters* **3**, 1183 (2003), ISSN 1530-6984.
- [78] S. Curtarolo, D. Morgan, K. Persson, J. Rodgers, and G. Ceder, *Physical Review Letters* **91** (2003), ISSN 0031-9007.
- [79] C. C. Fischer, K. J. Tibbetts, D. Morgan, and G. Ceder, *Nature Materials* **5**, 641 (2006), ISSN 1476-1122.
- [80] H. K. D. H. Bhadeshia, R. C. Dimitriu, S. Forsik, J. H. Pak, and J. H. Ryu, *Materials Science and Technology* **25**, 504 (2009), ISSN 0267-0836.
- [81] B. G. Sumpter and D. W. Noid, *Annual Review of Materials Science* **26**, 223 (1996), ISSN 0084-6600.
- [82] A. Abbasi, A. Dick, T. Hickel, and J. Neugebauer, *Acta Materialia* **59**, 3041 (2011).
- [83] G. Frommeyer and U. Bruex, *Steel Research International* **77**, 627 (2006).
- [84] Z. Pei, D. Ma, M. Friak, B. Svendsen, D. Raabe, and J. Neugebauer, *Physical Review B* **92** (2015).
- [85] K. Yang, W. Setyawan, S. Wang, M. Buongiorno Nardelli, and S. Curtarolo, *Nat Mater* **11**, 614 (2012), ISSN 1476-1122 (Print) 1476-1122 (Linking).
- [86] G. K. H. Madsen, *Journal of the American Chemical Society* **128**, 12140 (2006).
- [87] I. Opahle, G. K. Madsen, and R. Drautz, *Phys Chem Chem Phys* **14**, 16197 (2012), ISSN 1463-9084 (Electronic) 1463-9076 (Linking).
- [88] I. Opahle, A. Parma, E. J. McEniry, R. Drautz, and G. K. H. Madsen, *New Journal of Physics* **15** (2013), ISSN 1367-2630.
- [89] S. Kirklin, J. E. Saal, V. I. Hegde, and C. Wolverton, *Acta Materialia* **102**, 125 (2016).
- [90] M. W. Gaultois, T. D. Sparks, C. K. H. Borg, R. Seshadri, W. D. Bonificio, and D. R. Clarke, *Chemistry of Materials* **25**, 2911 (2013).
- [91] D. Lencer, M. Salinga, B. Grabowski, T. Hickel, J. Neugebauer, and M. Wuttig, *Nat Mater* **7**, 972 (2008).
- [92] T. Hammerschmidt, A. F. Bialon, D. G. Pettifor, and R. Drautz, *New Journal of Physics* **15** (2013), ISSN 1367-2630.
- [93] F. Kormann, D. Ma, D. D. Belyea, M. S. Lucas, C. W. Miller, B. Grabowski, and M. H. F. Sluiter, *Applied Physics Letters* **107** (2015).

- [94] H. Titrian, U. Aydin, M. Friak, D. Ma, D. Raabe, and J. Neugebauer, in *Symposium RR, Advanced Multiscale Materials Simulation, Toward Inverse Materials Computation* (2013), vol. 1524 of *MRS Proceedings*.
- [95] M. Friak, L.-F. Zhu, L. Lymperakis, H. Titrian, U. Aydin, A. M. Janus, H.-O. Fabritius, A. Ziegler, S. Nikolov, P. Hemzalova, et al., *Quantum-mechanical study of single-crystalline and polycrystalline elastic properties of Mg-substituted calcite crystals* (2014), vol. 592-593 of *Key Engineering Materials*, pp. 335–341.
- [96] L. F. Zhu, M. Friak, L. Lymperakis, H. Titrian, U. Aydin, A. M. Janus, H. O. Fabritius, A. Ziegler, S. Nikolov, P. Hemzalova, et al., *Journal of the Mechanical Behavior of Biomedical Materials* **20**, 296 (2013).
- [97] W. Kohn and L. J. Sham, *Physical Review* **140**, A1133 (1965).
- [98] J. P. Perdew, K. Burke, and M. Ernzerhof, *Physical Review Letters* **77**, 3865 (1996).
- [99] J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh, and C. Fiolhais, *Physical Review B* **46**, 6671 (1992).
- [100] A. D. Becke, *Physicak Review A* **38**, 3098 (1988).
- [101] N. Ashcroft and N. Mermin, *Solid State Physics* (Thomson Learning, New York 1976).
- [102] H. Hellmann, *The Journal of Chemical Physics* **3**, 61 (1935).
- [103] P. E. Blochl, *Physical Review B* **50**, 17953 (1994).
- [104] A. Glensk, B. Grabowski, T. Hickel, and J. Neugebauer, *Physical review x* **4** (2014), ISSN 2160-3308.
- [105] J. A. Puszkiel, P. Arneodo Larochette, and F. C. Gennari, *Journal of Alloys and Compounds* **463**, 134 (2008).
- [106] R. Nazarov, T. Hickel, and J. Neugebauer, *Physical Review B* **82**, 224104 (2010).
- [107] K. Reuter and M. Scheffler, *Physical Review B* **65**, 035406 (2001).
- [108] W. Frank, C. Elsässer, and M. Fähnle, *Physical Review Letters* **74**, 1791 (1995).
- [109] B. Grabowski, T. Hickel, and J. Neugebauer, *Physical Review B* **76**, 024309 (2007).
- [110] P. Ferrin, S. Kandoi, A. U. Nilekar, and M. Mavrikakis, *Surface Science* **606**, 679 (2012).
- [111] V. Ledentu, W. Dong, and P. Sautet, *Surface Science* **412-13**, 518 (1998).
- [112] K. Lee, M. Yuan, and J. Wilcox, *Journal of Physical Chemistry C* **119**, 19642 (2015).
- [113] T. I. Spiridonova, A. V. Bakulin, and S. E. Kulkova, *Physics of the Solid State* **57**, 1921 (2015).
- [114] R. Griessen, *Physical Review B* **38**, 3690 (1988).

- [115] R. Griessen and R. Feenstra, *Journal of Physics F: Metal Physics* **15**, 1013 (1985).
- [116] R. Griessen and A. Driessen, *Physical Review B* **30**, 4372 (1984).
- [117] K. Pearson, *Proceedings of the Royal Society London* **Iviii**, 240 (1895).
- [118] J. E. Jackson, *Varimax Rotation* (John Wiley & Sons, Ltd, 2014), ISBN 9781118445112.
- [119] U. Fayyad, G. PiatetskyShapiro, and P. Smyth, *Communications of the Acm* **39**, 27 (1996).
- [120] U. Fayyad, G. PiatetskyShapiro, and P. Smyth, *Ai Magazine* **17**, 37 (1996).
- [121] G. Kresse and J. Furthmüller, *Physical Review B* **54**, 11169 (1996).
- [122] S. Plimpton, *Journal of Computational Physics* **117**, 1 (1995).
- [123] A. Glensk, *Doktorarbeit* (Paderborn 2015).
- [124] M. Folk and E. Pourmal, *Hdf software process* (2004).
- [125] Postgresql, *Performance postgresql* (2015), URL <http://www.pytables.org/usersguide/optimization.html>.
- [126] J. Berkus, P. Eisentraut, M. Hagander, T. Lane, B. Momjian, and D. Page, *Postgresql*, www.postgresql.org (2015).
- [127] W. G. S. Microsystems), in *Proceedings of the 1st International Symposium on Cluster Computing and the Grid* (IEEE Computer Society, Washington, DC, USA, 2001), CCGRID '01, pp. 35–, ISBN 0-7695-1010-8.
- [128] F. Murnaghan, *National Academy of Sciences* **30**, 244 (1944).
- [129] K. Kopitzki and P. Herzog, *Einführung in die Festkörperphysik* (Teubner, Stuttgart 2002).
- [130] J. Friedel and C. M. Sayers, *Journal De Physique Lettres* **38**, L263 (1977).
- [131] D. M. P. Mingos, *Essential Trends in Inorganic Chemistry* (Oxford University Press, Oxford, 1998).
- [132] B. Cordero, V. Gomez, A. E. Platero-Prats, M. Reves, J. Echeverria, E. Cremades, F. Barragan, and S. Alvarez, *Dalton Transactions* pp. 2832–2838 (2008).
- [133] Experiments, *Bulk modulus* (2015), URL <http://www.webelements.com>.
- [134] M. Palumbo, B. Burton, A. C. E. Silva, B. Fultz, B. Grabowski, G. Grimvall, B. Hallstedt, O. Hellman, B. Lindahl, A. Schneider, et al., *Physica Status Solidi B-Basic Solid State Physics* **251**, 14 (2014), ISSN 0370-1972.
- [135] P. Janthon, S. J. Luo, S. M. Kozlov, F. Vines, J. Limtrakul, D. G. Truhlar, and F. Illas, *Journal of Chemical Theory and Computation* **10**, 3832 (2014), ISSN 1549-9618.

- [136] F. R. de Boer, R. Boom, W. C. M. Mattens, A. R. Miedema, and A. K. Niessen, *Cohesion in Metals: Transition Metal Alloys* (North-Holland, Amsterdam, 1988).
- [137] K. H. J. Buschow, P. C. P. Bouten, and A. R. Miedema, Reports on Progress in Physics **45**, 937 (1982), ISSN 0034-4885.
- [138] R. Griessen and R. Riesterer, *Heat of formation models* (Springer Berlin / Heidelberg, 1988).
- [139] E. Fromm and G. Hörz, Intern. Metals Rev. **15**, 269 (1980).
- [140] M. Nagasaka and T. Yamashina, Journal of the Less-Common Metals **45**, 53 (1976).
- [141] H. Behrens and G. Ebel, in Physica Data Series 5 (Fachinformationszentrum Energie, Physik, Mathematik, Karlsruhe).
- [142] A. Driessen, H. Hemmes, and R. Griessen, Z. Phys. Chem. **143**, 145 (1985).
- [143] C. Papastaikoudis, B. Lengeler, and W. Jager, Journal of Physics F: Metal Physics **13**, 2257 (1983).
- [144] E. Fromm and E. Gebhardt, *Gase und Kohlenstoff in Metallen* (Springer Berlin / Heidelberg, 1976).
- [145] R. Frank and J. Thomas Jr., Journal of Physics and Chemistry of Solids **16**, 144 (1960).
- [146] U. Aydin, L. Ismer, T. Hickel, and J. Neugebauer, Physical Review B **85** (2012), ISSN 1098-0121 1550-235X.
- [147] L. Ismer, T. Hickel, and J. Neugebauer, Physical Review B **81**, 094111 (2010).
- [148] Y. Syono, K. Kusaba, K. Fukuoka, Y. Fukai, and K. Watanabe, Physical Review B **29**, 6520 (1984).
- [149] B. Baranowski, S. Majchrzak, and T. B. Flanagan, Journal of Physics F: Metal Physics **1**, 258 (1971).
- [150] D. Tanguy, Y. Wang, and D. Connetable, Acta Materialia **78**, 135 (2014), ISSN 1359-6454.
- [151] A. C. Switendick, *The change in electronic properties on hydrogen alloying and hydride formation* (Springer Berlin / Heidelberg, 1978).